



**Magisterarbeit**  
zur Erlangung des akademischen Grades  
**Magister Artium Internationales Informationsmanagement**



## **„Helen“ – Die persönliche Informationsagentin im Fahrzeug**

-

**Usability-Evaluierung eines anthropomorphen Interface-Agenten in  
einem multimodalen Fahrerinformationssystem der Blaupunkt GmbH**

**Matthias Görtz**

(matthiasgoertz@gmx.de)

**Hildesheim im April 2006**

**Erstgutachterin: Prof. Dr. Christa Womser-Hacker**

**Zweitgutachter: Dr. Thomas Mandl**

**Betreuer der Blaupunkt GmbH: Dr. Luis Arévalo (CM-DI/ESH)**

## **Zusammenfassung**

Die vorliegende Masterarbeit untersucht den Mehrwert von anthropomorphen Interface-Agenten als Benutzerschnittstelle im Sprachdialog von Fahrerinformationssystemen. Vor diesem Hintergrund werden ein von der Blaupunkt GmbH erstellter Prototyp eines Sprachdialogsystems und ein um einen virtuellen Charakter der Charamel GmbH erweitertes System miteinander vergleichend evaluiert. Die dazu durchgeführte Experten-Evaluation und anschließende Benutzerstudie stellen in den Vordergrund der Untersuchung, inwiefern eine anthropomorphe Oberfläche die Akzeptanz der Anwender und die Usability des Sprachdialogs erhöht. Die Erkenntnisse dieser Untersuchung münden in eine Bewertung bezüglich des Mehrwerts virtueller Charaktere im gegebenen Anwendungskontext und in Richtlinien für die Gestaltung eines anthropomorphen Fahrerinformationssystems.

*Schlüsselwörter: Usability, Evaluierung, Benutzertest, Fahrerinformationssystem, Sprachdialog, Anthropomorphismus, virtuelle Charaktere, Avatare*

## **Abstract**

This M.A. Thesis deals with the added value of anthropomorphic agents as user interfaces in speech dialogues of driver information systems. In this matter a speech dialogue prototype of Blaupunkt GmbH and the system enhanced with a virtual character of Charamel GmbH are comparatively evaluated. Therefore an expert evaluation and a subsequent user study are carried out. The main focus of these examinations is the possible added value in terms of user acceptance and system usability of virtual characters in speech dialogue systems. The results of this study lead to a judgement on the added value of such user interfaces in the given context and guidelines for the design of anthropomorphic interfaces.

*Keywords: usability test, user acceptance, driver-information system, speech dialogue system, anthropomorphism, virtual characters, embodied conversational agents*

# Inhaltsverzeichnis

Abkürzungsverzeichnis.....	iv
Tabellenverzeichnis.....	iv
Abbildungsverzeichnis.....	v
<b>1. Einleitung.....</b>	<b>1</b>
1.1. Motivation.....	1
1.2. Aufgabenstellung und Ziel der Arbeit .....	2
1.3. Aufbau der Arbeit .....	3
<b>2. Usability von Informationssystemen im Fahrzeug.....</b>	<b>4</b>
2.1. Software-ergonomische Grundlagen.....	4
2.2. Methoden zur Evaluierung von Usability und Nutzerakzeptanz .....	6
2.2.1. Expertenorientierte Methoden.....	8
2.2.2. Benutzerorientierte Methoden.....	9
2.3. Bedienkonzepte von Fahrerinformationssystemen (FIS) .....	12
2.3.1. Bedienmodalitäten von FIS.....	13
2.3.2. Multimodale Fahrerinformationssysteme .....	15
2.3.3. Usability und Akzeptanz von Sprachdialogsystemen.....	18
<b>3. Anthropomorphe Interface-Agenten (AIA).....</b>	<b>21</b>
3.1. Begriffsabgrenzung .....	21
3.2. Bisherige Einsatzbereiche virtueller Charaktere .....	23
3.3. Potenziale für den Sprachdialog im Fahrerinformationssystem .....	24
<b>4. Usability-Evaluierung eines AIA im multimodalen FIS .....</b>	<b>28</b>
4.1. Beschreibung der Prototypen.....	29
4.1.1. SDS Demonstrator.....	29
4.1.2. „Helen“ – Die persönliche Informationsagentin .....	33
4.2. Methodenwahl.....	36
4.3. Heuristische Evaluierung der Prototypen .....	37
4.3.1. Konzeption und Durchführung der Experten-Evaluation .....	37

4.3.2.	Ergebnisse .....	40
4.3.3.	Schlussfolgerungen der heuristischen Evaluierung .....	46
4.3.3.1.	Anpassung der Prototypen .....	47
4.3.3.2.	Hypothesen für den Benutzertest .....	48
4.3.3.3.	Konzeption der Aufgabenstellung .....	49
4.4.	Vergleichender Benutzertest der Prototypen .....	51
4.4.1.	Probandenauswahl .....	52
4.4.2.	Aufbau und Durchführung .....	56
4.4.3.	Ergebnisse .....	66
4.4.3.1.	Leistungsdaten (Task Performance Data) .....	66
4.4.3.2.	Post-Task-Fragebogen .....	72
4.4.3.3.	Physiologische Daten .....	81
<b>5.</b>	<b>Auswertung und Interpretation der Ergebnisse .....</b>	<b>86</b>
5.1.	Überprüfung der Hypothesen .....	86
5.2.	Ursachenanalyse der beobachteten Ergebnisse .....	89
5.3.	Schlussfolgerungen für den Einsatz anthropomorpher Interface-Agenten ..	92
<b>6.</b>	<b>Fazit und Ausblick .....</b>	<b>95</b>
	<b>Literaturverzeichnis .....</b>	<b>97</b>
	<b>Eigenständigkeitserklärung nach §26 Abs. 6 der Magisterprüfungsordnung .</b>	<b>102</b>

## **Anhang**

Anhang A – Usability-Heuristiken  
Anhang B – Mängelliste der Experten-Evaluierung  
Anhang C – Aufgabenszenarien des Benutzertests  
Anhang D – Ergebnisse der Fragebögen  
Anhang E – Auswertung physiologischer Kennzahlen

## **CD-ROM**

Magisterarbeit in pdf-Format  
Fragebogen der Experten-Evaluierung  
Benutzertest: Erfassungsvorlage für den Versuchsleiter  
Post-Task-Fragebogen  
Beispielvideo des Versuchablaufs (mpeg)

## Abkürzungsverzeichnis

<i>Abkürzung</i>	<i>Begriff</i>
AIA	Anthropomorpher Interface-Agent
BPM	Beats Per Minute (Herzfrequenz)
DIN	Deutsches Institut für Normung
EDA	Elektrodermale Aktivität
EXP	Technische Vorerfahrung der Probanden
FIS	Fahrerinformationssysteme
GUI	Graphical User Interface
ISO	International Organization of Standardization
PTT	Push-To-Talk
SDS	Speech Dialogue System
TCT	Task Completion Time
TCR	Task Completion Rate
TTS	Text-To-Speech

## Tabellenverzeichnis

Tabelle 4.1: Severity Rating der Experten-Evaluation .....	40
Tabelle 4.2: Aufschlüsselung der Probandenauswahl.....	55
Tabelle 4.3: Item 4 – Teil 1 / Schwierigkeiten der Benutzer .....	77
Tabelle 4.4 Vergleich der physiologischen Kennzahlen Pbn008.....	84

## Abbildungsverzeichnis

Abbildung 2.1: LUCID (vgl. Kreitzberg 1996. nach Shneiderman 1998:104).....	7
Abbildung 2.2: Travel Pilot Rome. Bedienung im Dreh-/Drück-Konzept.....	13
Abbildung 2.3: Konzept eines multimodalen Kfz-Informationssystems .....	17
Abbildung 3.1: Virtuelle Charaktere der Charamel GmbH.....	22
Abbildung 4.1: SDS Demonstrator .....	29
Abbildung 4.2: SDS Help Popup bei aktiver Spracheingabe .....	31
Abbildung 4.3: Bediensequenz „Abspielen CD 3“.....	32
Abbildung 4.4: Prototyp <i>Helen</i> .....	34
Abbildung 4.5: Beispiele der Körpersprache des AIA <i>Helen</i> .....	35
Abbildung 4.6: Verteilung der notierten Mängel auf die heuristischen Kategorien.....	41
Abbildung 4.7: Severity-Rating der Experten-Evaluation.....	42
Abbildung 4.8: Fragebogenergebnisse der Experten: Usability im Vergleich .....	45
Abbildung 4.9: Angepasste Prototypen für den Benutzertest .....	47
Abbildung 4.10: Zielgruppeneinteilung der Blaupunkt GmbH.....	53
Abbildung 4.11: Screening Internetseite ( <a href="http://www.neugier-innovation.de">www.neugier-innovation.de</a> ) .....	54
Abbildung 4.12: Raumaufteilung des Usability-Labors .....	58
Abbildung 4.13: Versuchsaufbau mit Proband .....	59
Abbildung 4.14: Phasische EDA bei Frustration des Probanden.....	61
Abbildung 4.15: Biofeedbacksensoren zur Messung physiologischer Daten .....	63
Abbildung 4.16: Überblick Task 1 Completion Time.....	67
Abbildung 4.17: Überblick Task 1 Completion Rate .....	69
Abbildung 4.18: Überblick Task 2 Completion Time.....	70
Abbildung 4.19: Fragebogen Item Usability.....	73
Abbildung 4.20: Fragebogen Item Emotion .....	74
Abbildung 4.21: Überblick Gesamtbewertung Items 1 - 3 .....	75
Abbildung 4.22: Item 4 - Teil 2 / Hilfefunktionen im Vergleich .....	78
Abbildung 4.23: Präferenz der Systeme nach Merkmalsunterscheidungen .....	80
Abbildung 4.24: EDA Aufzeichnung Proband 042 Task 1 SDS.....	82
Abbildung 4.25: Physiologische Kennzahlen Pbn008 im Vergleich.....	83

# 1. Einleitung

*"Avoid presenting computers as people."* (SHNEIDERMAN 1998: 385)

*"Embodied conversational agents provide  
a promising option for presenting information to users."*

(RIST et al.. 2004: 377)

## 1.1. Motivation

In seinem alltäglichen Leben stehen dem Menschen in jeder Situation nützliche Werkzeuge zur Verfügung, die ihn dabei unterstützen, seine Arbeit zu vollbringen und seine Ziele einfacher und schneller zu erreichen. Im Laufe der Zeit wurden diese Werkzeuge von handwerklichen Gegenständen zu elektronischen Anwendungen weiterentwickelt. Die Art der Steuerung dieser Hilfsmittel mag sich grundlegend geändert haben, doch schon immer war dabei das Ziel, die Bedienung von Werkzeugen so intuitiv wie möglich zu gestalten. *Intuition* ist „das unmittelbare, nicht diskursive, nicht auf Reflexion beruhende Erkennen, Erfassen eines Sachverhalts oder eines komplizierten Vorgangs“. (DUDEN 2001: 457) Und da dem Menschen nichts unmittelbarer ist als die Interaktion mit anderen Menschen, gibt es viele Stimmen wie auch Rist et al., die vermuten, dass die Angleichung der Mensch-Maschine-Interaktion an den zwischenmenschlichen Dialog die optimale Gestaltung einer Mensch-Maschine-Schnittstelle darstelle. Doch gleichermaßen gibt es jene, die ähnlich Shneiderman behaupten, dass gerade eine solche Gleichsetzung irritierend und hinderlich für die erfolgreiche Bedienung einer Software-Anwendung sei. So ist in der Wissenschaft ein intensiver Diskurs zu beobachten, der für jeweilige Anwendungssituationen untersucht, welche Auswirkungen der Einsatz einer anthropomorphen, also menschenähnlichen, Bedienoberfläche auf die Gebrauchstauglichkeit und Nutzerakzeptanz eines Systems hat.

## **1.2. Aufgabenstellung und Ziel der Arbeit**

Auch die Verantwortlichen der *Blaupunkt GmbH* stellten sich die Frage nach den Potenzialen eines virtuellen Charakters als Schnittstelle zu ihrem Fahrerinformationssystem. Bereits in der Vergangenheit war das Navigationssystem der Firma mit einer persönlichen Stimme versehen worden, die einen natürlichen Klang und einen hohen Wiedererkennungswert der *Blaupunkt*-Produkte garantierte. Diese Stimme war auch Bestandteil der weiteren Entwicklung von Sprachdialogsystemen in den Produkten der nächsten Generation. Solche Entwicklungsprojekte brachten die obige Diskussion auf die Tagesordnung und es wurde der *Charamel GmbH* der Auftrag gegeben, zu Versuchszwecken den Prototyp eines Sprachdialogsystems mit einem menschlichen Antlitz zu versehen. So entstand „Helen“, die persönliche Informationsagentin im Fahrzeug und zudem der Auftrag für diese Magisterarbeit.

Ziel dieser Arbeit ist es, für die spezielle Anwendungssituation im Fahrzeug zu evaluieren, welcher Mehrwert dem Einsatz eines anthropomorphen Interface-Agenten in Bezug auf Gebrauchstauglichkeit und Benutzerakzeptanz eines multimodalen Fahrerinformationssystems zukommt. Steigert ein virtueller Charakter die Akzeptanz der Sprachbedienung bzw. die Attraktivität des Produkts? Können durch die sprachliche Interaktion mit einer virtuellen Informationsagentin Eingabefehler vermieden oder schneller korrigiert werden? Wirkt sich der Einsatz von „Helen“ positiv auf die allgemeine Benutzerzufriedenheit aus? Oder sorgt gerade vielleicht die Anwendungssituation dafür, dass der Benutzer<sup>1</sup> ablehnend auf ein solches System reagiert?

Die aus dieser Arbeit resultierenden Ergebnisse sollen helfen, diese Fragen zu beantworten und Empfehlungen auszusprechen, inwiefern weitere Entwicklungen von multimodalen Fahrerinformationssystemen mit virtuellen Charakteren sinnvoll sind und welche Richtlinien sich für die Gestaltung eines Sprachdialogsystems in möglicher Kombination mit einem Interface-Agenten ableiten lassen.

---

<sup>1</sup> In dieser Arbeit wird zur besseren Lesbarkeit nur die männliche Form verwendet, wenn in einem allgemeinen Sinne von Personen die Rede ist.



### **1.3. Aufbau der Arbeit**

Vor dem Hintergrund der eingangs dargelegten Motivation und Zielsetzung wurde im Rahmen dieser Masterarbeit eine zweistufige, vergleichende Usability-Evaluierung durchgeführt. Die Ausarbeitung besteht aus sechs Kapiteln.

In *Kapitel 2* werden die Grundlagen zum Verständnis und zur Bearbeitung der Ausgangsfragestellungen erläutert. Dies beinhaltet zunächst die Definition grundlegender Prinzipien zur Gestaltung von Dialogsystemen. Dazu gehört auch die Beschreibung methodischer Grundlagen zum Messen der Usability eines Mensch-Maschine-Systems. Die Betrachtung der Fahrzeugdomäne und den damit verbundenen Geräten und Funktionalitäten führen zu der Beschreibung der unterschiedlichen Bedienkonzepte im Auto. In diesem Zusammenhang wird die Bedeutung der Multimodalität erläutert. Die grundlegenden Usability-Richtlinien werden daraufhin für die Konzeption sprachbedienbarer Fahrerinformationssysteme erweitert. Die Betrachtung der besonderen Schwierigkeiten der Anwender in der Interaktion mit Sprachdialogsystemen führt dann zu der Darstellung der Technologie der anthropomorphen Interface-Agenten. *Kapitel 3* klärt Begrifflichkeiten, zeigt beispielhafte Einsatzbereiche auf und formuliert erste Thesen für den Einsatz solcher virtuellen Charaktere in Fahrerinformationssystemen. Das Kernelement der Arbeit stellt *Kapitel 4* dar. In diesem Abschnitt wird die Usability-Evaluierung zweier prototypischer Sprachdialogsysteme vorgestellt. Dafür werden die beiden Prototypen beschrieben, die Wahl der Methoden der Untersuchung dokumentiert und schließlich die Evaluierungsschritte und deren Ergebnisse präsentiert. Die Untersuchung ist dabei zweistufig aufgebaut. Die Grundlage bildet eine Experten-Evaluierung, deren Vorgehen und Ergebnisse in *Kapitel 4.3* erläutert werden. Die Schlussfolgerungen dieser Untersuchung fließen dann in einen Benutzertest zur Usability und Akzeptanz eines anthropomorphen Interface-Agenten im Sprachdialog eines Fahrerinformationssystems ein. Nach der Ergebnispräsentation folgt in *Kapitel 5* die Interpretation und Auswertung der in den Evaluierungen gewonnenen Daten. Die vorher formulierten Hypothesen werden geprüft und die Aussagekraft der gewonnenen Erkenntnisse bewertet. Das abschließende *Kapitel 6* dient der Zusammenfassung der Untersuchung und einem Ausblick auf weitere Entwicklungen.

## **2. Usability von Informationssystemen im Fahrzeug**

Das folgende Kapitel stellt die notwendige Grundlage für die Konzeption und Durchführung einer Usability-Evaluierung dar. Es definiert die grundlegenden Begriffe und Methoden der Software-Ergonomie, bevor es auf den speziellen Anwendungskontext im Fahrzeug eingeht. Die Beschreibung der Bedienkonzepte von Fahrerinformationssystemen hebt die Bedeutung multimodaler Schnittstellen gerade im Fahrzeug heraus. Die anschließenden erweiterten Richtlinien zur Gestaltung von Sprachdialogsystemen gehen auf die speziellen Anforderungen solcher Schnittstellen ein und bilden so die Basis, um später ein Urteil fällen zu können, ob das untersuchte System vom Anwender akzeptiert und entsprechend bedient werden kann.

### **2.1. Software-ergonomische Grundlagen**

Der Gestaltung der Benutzeroberfläche eines Fahrerinformationssystems liegen die Erkenntnisse der Software-Ergonomie zugrunde. Die traditionelle Disziplin der Ergonomie (gr. *érgon* = *Arbeit, Werk, Tat*) ist die „Wissenschaft von den Leistungsmöglichkeiten u. –grenzen des arbeitenden Menschen sowie der besten wechselseitigen Anpassung zwischen dem Menschen u. seinen Arbeitsbedingungen“. (DUDEN 2001: 278) Die ursprüngliche arbeitswissenschaftliche Konzeption von ergonomischen Gebrauchsgegenständen wurde im Laufe der wachsenden Bedeutung der Informationstechnologie um die Disziplin der Software-Ergonomie erweitert, welche die Schnittstelle von Mensch und Computer betrachtet. „Die Software-Ergonomie hat das Ziel, die Software eines Computers, mit der der Benutzer arbeitet, an die Eigenschaften und Bedürfnisse dieser Benutzer anzupassen, um ihnen einen hohen Nutzen möglichst vieler relevanter Fähigkeiten und Fertigkeiten zu ermöglichen.“ (BALZERT et al. 1999: 655) Somit werden die ergonomischen Grundsätze über handwerkliche Werkzeuge hinaus auf die Konzeption von Computerprogrammen angewendet.

Gegenstand der Software-Ergonomie ist die *Usability*. Dieses Kompositum aus dem englischen *to use* (benutzen) und *the ability* (die Fähigkeit) wurde ursprünglich übersetzt als Benutzbarkeit oder Gebrauchstauglichkeit, wird heute aber zunehmend auch in der deutschsprachigen Literatur verwendet. Die Internationale Organisation für Standardisierung (ISO) definiert diesen Begriff in der ISO-Norm 9241-11 wie folgt: Gebrauchstauglichkeit ist „das Ausmaß, in dem ein Produkt durch *bestimmte* Benutzer in einem *bestimmten* Kontext genutzt werden kann, um *bestimmte* Ziele *effektiv, effizient* und *zufrieden stellend* zu erreichen“. (DIN 2004: 96) Somit gilt es bei der Frage nach der Usability eines Systems folgende Fragen zu beantworten:

*Effektivität* – Kann der Benutzer mit dem Produkt sein formuliertes Ziel erreichen?

*Effizienz* – Wie viel Aufwand und Zeit kostet ihn die Zielerreichung?

*Zufriedenheit* – Wie zufrieden ist der Benutzer mit der Benutzung des Systems?

Diese Fragen dienen als Kriterien einer Usability-Evaluierung. Allerdings sind für eine umfassende Bewertung dieser Parameter eine Berücksichtigung der Anwendungssituation (s. *Kapitel 2.3.*) und zunächst eine Konkretisierung dieser Leitfragen nötig. Wie muss eine Interaktion zwischen Mensch und Maschine gestaltet sein, um die oben genannten Kriterien zu erfüllen? Teil 10 der ISO-Norm 9241 bietet mit den „Richtlinien der Dialoggestaltung“ einen ersten Ansatz. Es werden folgende Systemeigenschaften definiert:

- *Aufgabenangemessenheit*

„Ein Dialog ist aufgabenangemessen, wenn er den Benutzer unterstützt, seine Arbeitsaufgabe effektiv und effizient zu erledigen.“

- *Selbstbeschreibungsfähigkeit*

„Ein Dialog ist selbstbeschreibungsfähig, wenn jeder einzelne Dialogschritt durch Rückmeldung des Dialogsystems unmittelbar verständlich ist oder dem Benutzer auf Anfrage erklärt wird.“

- *Steuerbarkeit*

„Ein Dialog ist steuerbar, wenn der Benutzer in der Lage ist, den Dialogablauf zu starten sowie seine Richtung und Geschwindigkeit zu beeinflussen, bis das Ziel erreicht ist.“

- *Erwartungskonformität*

„Ein Dialog ist erwartungskonform, wenn er konsistent ist und den Merkmalen des Benutzers entspricht, z.B. seinen Kenntnissen aus dem Arbeitsgebiet, seiner Ausbildung und seiner Erfahrung sowie den allgemein anerkannten Konventionen.“

- ***Fehlertoleranz***

„Ein Dialog ist fehlertolerant, wenn das beabsichtigte Arbeitsergebnis trotz erkennbar fehlerhafter Eingaben entweder mit keinem oder mit minimalem Korrekturaufwand seitens des Benutzers erreicht werden kann.“

- ***Individualisierbarkeit***

„Ein Dialog ist individualisierbar, wenn das Dialogsystem Anpassungen an die Erfordernisse der Arbeitsaufgabe sowie an die individuellen Fähigkeiten und Vorlieben des Benutzers zulässt.“

- ***Lernförderlichkeit***

„Ein Dialog ist lernförderlich, wenn er den Benutzer beim Erlernen des Dialogsystems unterstützt und anleitet.“

(DIN 2004: 85ff.)

Diese grundsätzlichen Regeln zur Gestaltung der Schnittstelle zwischen Mensch und Computersystem bilden auch die Basis für das benutzerorientierte Design eines Fahrerinformationssystems. Dabei gelten die international anerkannten Normen zwar als Richtlinien, in der Evaluierung der vorliegenden Systeme kommt es aber besonders auf Anwendungsmodalität und –kontext an. Gerade im Fahrzeug ist ein spezielles Nutzungsumfeld gegeben, das sensibel ist und besondere Anforderungen an Informationssysteme stellt. Die speziellen Modalitäten und Funktionalitäten eines Fahrerinformationssystems werden im *Abschnitt 2.3.* beschrieben und münden in eine Erweiterung der hier vorgestellten allgemeinen Usability-Richtlinien. Der direkt anschließende Teil soll zunächst einmal in Ergänzung zu den software-ergonomischen Grundlagen die möglichen Untersuchungsmethoden für die Prüfung auf Usability und Benutzerakzeptanz vorstellen.

## ***2.2. Methoden zur Evaluierung von Usability und Nutzerakzeptanz***

Im Bereich der Software-Ergonomie bzw. der Usability-Forschung wurden vielfältige Methoden entwickelt, um zu untersuchen, inwiefern eine Webseite, eine Software-Anwendung bzw. ein Bildschirmdialogsystem benutzergerecht gestaltet sind. Die Erkenntnis, dass eine solche Untersuchung grundlegend zum Erfolg eines Produktes beiträgt, ist spät gereift, hat sich jedoch letztlich durchgesetzt. Shneiderman begründete dies so: „Careful attention to user-centred design issues at the early stages of software development has been shown to reduce both development time

and cost dramatically“. (1998: 104) Im Ansatz der Logical User-Centred Interactive Design Methodology (LUCID) von Kreitzberg wird deutlich, wie die Evaluierung der Usability eines Produkts in den Entwicklungsprozess eingebunden sein kann. (s. Abb. 2.1)

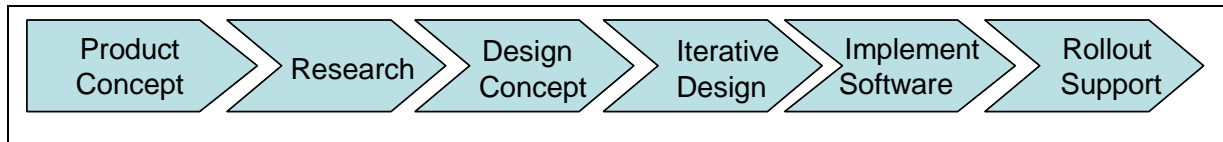


Abbildung 2.1: Logical User-Centred Interactive Design Methodology<sup>2</sup>

Dieses Modell sieht vor, dass die Usability-Evaluierung ein fester Bestandteil der Produktentwicklung ist und eingebunden in einen iterativen Prozess schon während der Entwicklung einer Software Erkenntnisse zur Gebrauchstauglichkeit des Produkts liefert. Die vierte Phase ist dabei wie folgt definiert:

**„Stage 4: Do iterative design and refinement:**

*Expand key-screen prototype into full system.*

*Conduct heuristic and expert reviews.*

*Conduct full-scale usability tests.*

*Deliver prototype and specification.”*

(Kreitzberg 1996 nach Shneiderman 1998: 105)

Je nach Entwicklungsstand der Prototypen können diverse Methoden eingesetzt werden, die vom einfachen *Debugging*<sup>3</sup> bis zu umfassenden Erhebungen reichen. Hierbei ist zunächst in experten- und benutzerorientierte Untersuchungen zu unterscheiden. Expertenorientierte Methoden können schon in einem frühen Entwicklungsstadium eingesetzt werden, da die professionelle Analyse sich auf einzelne Aspekte eines Systems beschränken kann. Sie konzentrieren sich auf die Einhaltung der Usability-Richtlinien. Benutzerorientierte Erhebungen basieren auf umfangreichen Produkt- und Benutzertests mit messbaren Parametern. Sie bieten

---

<sup>2</sup> Die Abbildung der Phasen der LUCID-Methode stellt einen schematischen Überblick über Kreitzbergs Ausführungen bei Shneiderman (1998: 105) dar.

<sup>3</sup> (engl.) = (EDV) Vorgang bei der Programmherstellung, bei dem das Programm getestet wird und die entdeckten Fehler beseitigt werden. (DUDEN 2001: 198)

neben der Erkenntnis zu Effektivität und Effizienz einer Bedienoberfläche vor allem Aufschluss über die Nutzerakzeptanz des verwendeten Designs.

### 2.2.1. Expertenorientierte Methoden

Die Evaluierung eines Systems durch Experten bietet zunächst offensichtliche Vorteile: Professionelle Analysen eines Prototypen können innerhalb kurzer Zeit und in enger Zusammenarbeit mit dem Entwicklungsteam erstellt werden. Dabei kann die auf Erfahrung und internationalen Normen basierende Einschätzung der Experten schon früh in der Entwicklungsphase mögliche Problemstellen identifizieren. Die Ergebnisse einer Experten-Evaluierung können ein formaler Bericht oder Änderungsempfehlungen sein. Für die Vorgehensweise einer solchen Untersuchung stehen diverse Methoden zur Verfügung: Heuristische Evaluierung, Richtlinien-Untersuchung, Konsistenz-Inspektion, formelle Usability-Inspektion, etc.. (vgl. SHNEIDERMAN 1998: 126) Sie unterscheiden sich in der Umsetzung leicht, stimmen aber allgemein darin über ein, dass sie jeweils eine Benutzerschnittstelle auf Einhaltung bzw. Abweichungen eines definierten Ideals untersuchen. In einer solchen Analyse wird neben den objektiven Betrachtungen häufig noch die Perspektive der möglichen Anwender eines Systems betrachtet. In einem *Cognitive Walkthrough* versetzen sich die Experten in die Anwendungssituation und simulieren Bedienschritte aus Benutzersicht. (ebd.) Die Kombination beider Perspektiven ermöglicht, neben der reinen Betrachtung von allgemeinen Richtlinien zu analysieren, inwiefern ein Anwender diese Aspekte einer Schnittstelle wahrnimmt und mögliche Bedienfehler entstehen könnten.

Als ein Beispiel der oben genannten Methoden soll hier die heuristische Evaluierung näher erläutert werden. „Der Begriff *heuristisch* beschreibt das methodische Vorgehen zur Identifikation von Problemen.“ (PREIM 1999: 242) In Anwendung auf die Usability eines Systems, wird meist von mehreren Experten geprüft, inwieweit eine Benutzerschnittstelle mit anerkannten Prinzipien der Usability, den so genannten Heuristiken, übereinstimmt. (vgl. SCHWEIBENZ / THISSEN 2002: 100) Diese Prinzipien formulieren Optimalzustände und lassen gewissen Raum für Interpretation.

Der Experte sollte demnach mit den Heuristiken vertraut sein und kann diese dann auf das Gesamtsystem aber auch auf einzelne Aspekte der Benutzerschnittstelle anwenden. Anhand des in den Heuristiken beschriebenen Ideals kann der Experte ein System also detailliert auf Versäumnisse bezüglich gewisser Kategorien untersuchen. Diese Art der kontrollierten Analyse verhindert, dass Einzelpersonen einen Gesamteindruck eines Systems wiedergeben, der zwar auf fachlicher Expertise aber eventuell auf keinem einheitlichen Usability-Verständnis basiert. Eine Vorgabe diverser Kategorien und formulierter Idealzustände schafft ein durchgängiges Bewertungsmuster, anhand dessen jeder Experte einzeln und später in der Diskussion gemeinsam die Usability des zu untersuchenden Systems bewertet. Diese Methode hat also generell zum Ziel, dass eine analytische Untersuchung die eventuellen Missstände oder Mängel einer Benutzerschnittstelle erfasst. Hierbei weist Shneiderman zusätzlich darauf hin: "The reviewer notes possible problems for discussion with the designers, but solutions generally should be left for the designers to produce." (1998: 126) Die Rollenverteilung in dieser Evaluationsphase ist also klar definiert. Denn ein Experte, der aufgrund seiner Erfahrung und anhand der Richtlinien auf Usability-Mängel eines Systems hinweist, kann und muss nicht unbedingt die Lösung dieser Probleme bieten.

Bei einer solchen Vorgehensweise ist ebenso zu beachten, dass der Expertenblick auch einige Aspekte unbetrachtet lässt. So fällt es Experten schwer einzuschätzen, wie sich der typische Benutzer oder besonders ein Erstnutzer verhalten wird. Die Usability-Richtlinien formulieren zwar die optimalen Eigenschaften einer Benutzerschnittstelle, geben jedoch keine Antwort auf Nutzerpräferenzen und –akzeptanz. Insofern ist es unbedingt notwendig, die späteren Nutzer des Systems ebenfalls in die Konzeption mit einzubeziehen. Hierfür stehen einige benutzerorientierte Methoden zur Verfügung.

### **2.2.2. Benutzerorientierte Methoden**

Evaluationsmethoden, die den potenziellen Benutzer in die Bewertung einer Benutzerschnittstelle miteinbeziehen, sind bei der Entwicklung eines interaktiven

Systems von großer Bedeutung. Sie betrachten die Erfahrungen des Nutzers im Umgang mit einem Produkt. „Bei der empirischen (erfahrungsbezogenen) Evaluierung wird ein Prototyp mit Endbenutzern getestet. [...] Generell werden diese Verfahren aber eher in späten Phasen, z.B. nach der Auswertung und Problembeseitigung, die eine erste heuristische Evaluierung erforderlich gemacht hat, angewendet.“ (PREIM 1999: 245) Ist im Vorfeld bereits noch so viel analytisches Expertenwissen in die Konzeption eingeflossen, so entscheiden am Ende doch die empirischen Erkenntnisse über Akzeptanz und Erfahrungen der Nutzer über ein wirklich nutzerfreundliches Bedienkonzept.

Bei der Miteinbeziehung der Benutzer in die Evaluation einer Schnittstelle ist zu unterscheiden in Methoden der *Benutzerbeteiligung* und der *Benutzerbeobachtung*. Damit sind zum einen die Methoden gemeint, „die mit Befragungen der Benutzer verbunden sind, wie beispielsweise Benutzerbefragung mit Fragebögen, Interviews und Fokusgruppen-Interviews mit Benutzern. [Sie] dokumentieren Aussagen über potenzielles oder tatsächliches Verhalten. Dagegen dokumentieren mit Beobachtung verbundene Methoden, beispielsweise Produkttests im Labor, tatsächliches Verhalten in einer konkreten Situation.“ (SCHWEIBENZ / THISSEN 2002: 118ff.) Es ist also bei der Messung benutzerspezifischer Daten wichtig, zwischen einer gewissen Meinung bzw. Präferenz und tatsächlichen, objektiven Daten zu unterscheiden.

Die Beobachtung von Benutzern in der Interaktion mit einem System lässt sich in einem Usability-Test realisieren. In einem Labor lässt man dabei Probanden mit dem Prototypen eines Systems definierte Aufgabenstellungen bearbeiten und dokumentiert Leistungsdaten und Beobachtungen. „Je nach Ausstattung [des Labors] besteht die Möglichkeit, bestimmte Tätigkeiten am Computer zu messen (z.B. Arbeits-/Leerlaufzeiten, Zustände, Eingabegeschwindigkeiten) und diese getrennt oder per Video aufzuzeichnen.“ (PREIM 1999: 250) Zur Erhebung der Daten steht eine Vielzahl an Methoden zur Verfügung. So gibt es Tests, die ein späteres System simulieren (*Wizard-of-Oz*) oder den Nutzer zum *lauten Denken* anregen, um sein Vorgehen nachzuvollziehen, genau so wie eine *Logfile-Analyse* zur Aufzeichnung der Nutzeraktionen und *Videoaufzeichnungen*. (vgl. SHNEIDERMAN 1998: 130ff.) Je nach Anwendungsszenario und Untersuchungsgegenstand gilt es die passende Methode zu wählen und die für die Untersuchung relevanten Parameter zu definieren.



Eine abschließende statistische Auswertung kann somit Erkenntnisse zur objektiven Usability eines Systems treffen, indem das tatsächliche Verhalten von potenziellen Anwendern dokumentiert und statistisch ausgewertet wird. Eine lange Bearbeitungszeit oder auch eine hohe Anzahl an Bedienfehlern und Eingabeschritten weist auf reale Nutzungsprobleme und demnach Usability-Mängel hin. Die Aussagekraft der ausgewerteten Daten ist jedoch immer ins Verhältnis zu der vorhandenen Probandenauswahl und der Testumgebung zu setzen.

Als wichtiges Element der Usability gilt es ebenso die subjektive Bewertung eines Systems zu evaluieren. „Written user surveys are a familiar, inexpensive, and generally acceptable companion for usability tests and expert reviews“. (a.a.O.: 132) So bieten sich gerade in Kombination mit einem Benutzertest anschließende Fragebögen an, um den Probanden die Möglichkeit zu geben, subjektive Eindrücke in Bewertungsskalen oder freien Bemerkungen zum Ausdruck zu bringen. Shneiderman entwickelte so beispielsweise den *Questionnaire for User Interaction Satisfaction (QUIS)*, der neben Bewertungen von Systemelementen wie Darstellung, Terminologie und Systemstatus, Erlernbarkeit, etc. auch die emotionale Reaktion des Nutzers abfragt. (vgl. a.a.O.: 136) Neben der Evaluation im Anschluss an einen Benutzertest gibt es noch weitere subjektive Methoden wie eine Befragung während der Interaktion mit einem Prototyp, Interviews in Fokusgruppen oder auch Leitfadeninterviews in Einzelgesprächen. Am Ende einer Untersuchungsmethode mit Benutzerbeteiligung soll so neben den empirischen Daten aus der Beobachtung ein differenziertes Meinungsbild zu Akzeptanz und Bewertung eines Systems seitens potenzieller Kunden bzw. Anwender entstehen. Dabei muss die Gegenüberstellung messbarer und subjektiver Ergebnisse nicht unbedingt ein einheitliches Bild ergeben. In Kombination mit den analytischen Erkenntnissen der expertenorientierten Methoden lassen sich so aber die Usability und die Akzeptanz eines Dialogsystems umfassend evaluieren.

Nach Betrachtung des Usability-Begriffs und dessen Messbarkeit bringt das folgende Kapitel diesen Sachverhalt in Verbindung mit dem Anwendungskontext der Informationssysteme der Blaupunkt GmbH. Die Betrachtung der domänenspezifischen Systemumgebung ist notwendig, um die allgemeinen Richtlinien zur Bedienbarkeit anzupassen und zu spezifizieren.

### **2.3. Bedienkonzepte von Fahrerinformationssystemen (FIS)**

Die Mensch-Maschine-Schnittstelle im Fahrzeug ist zunächst definiert durch die elementaren Instrumente im Inneren, die der Steuerung des Automobils dienen. Die Primäraufgabe des Fahrers besteht darin, mit Lenkrad, Schalthebel, Blinker und anderen Schaltern und Hebeln das Fahrzeug zu steuern. Hierzu werden ihm die notwendigen Informationen, wie Geschwindigkeit, Tankfülle und weitere für die Primäraufgabe essentielle Daten angezeigt. Im Laufe der Erweiterung elektronischer Funktionalitäten im Fahrzeug kamen Sekundäraufgaben rund um die Bedienung des Autoradios, des CD-Spielers, sowie auch des Navigationssystems und Autotelefon hinzu. Sie stellen neue Komfortfunktionalitäten dar. Zudem wuchs auch die Anzahl elektronischer Hilfen zur Steuerung des Fahrzeugs. Die daraus entstandene Informationsfülle droht den Fahrer bei seiner Primäraufgabe mehr zu stören als zu unterstützen. So ist es von enormer Bedeutung, die Bedienschnittstelle zu diesen Funktionen benutzergerecht zu gestalten und den Aufwand der Bedienung sekundärer Aufgaben zu minimieren.

Zunächst gilt es zu trennen: Zum einen handelt es sich um Anwendungen, die die Sicherheit des Fahrers und die Handhabbarkeit des Fahrzeugs erhöhen sollen. Elektronische Hilfen wie das Anti-Blockier-System, Elektronische-Stabilitäts-Programm oder gar Nachtsichtassistentz werden zusammengefasst unter dem Begriff Fahrerassistenzsystem (FAS). Ein solches System wirkt sich meist ohne direkte Steuerung des Fahrers unterstützend auf dessen Primäraufgabe aus. Zum anderen stehen im Fahrzeug die alltäglichen Informations- und Kommunikationstechnologien zur Verfügung. Diese ermöglichen neben der Steuerung des Fahrzeugs Sekundäraufgaben wie das Hören von Musik, das Telefonieren oder auch ein gezieltes Navigieren zu einer Adresse. Der Benutzer muss diese Anwendungen direkt steuern, da sie von seiner Eingabe abhängige Aufgaben ausführen. Diese unter dem Begriff Fahrerinformationssystem (FIS) zusammengefassten Funktionalitäten sind Gegenstand dieser Usability- und Akzeptanzuntersuchung.

### 2.3.1. Bedienmodalitäten von FIS

Die Anzahl der Informations- und Kommunikationsfunktionen im Fahrzeug ist historisch gewachsen. So ist das Radio schon seit 1938 ein fester Bestandteil des Automobils, während erste CD-Spieler oder Navigationsgeräte Anfang der 80er Jahre Einzug ins Cockpit erhielten. Zunächst handelte es sich also primär um Einzelgeräte, die entweder in die Erstausrüstung eines Neuwagens integriert waren oder später zusätzlich zugekauft und nachgerüstet wurden. Diese Geräte wurden bisher in der Mittelkonsole meist in einem 1-DIN Schacht installiert. Um das Autoradio zum Beispiel einzuschalten oder den Sender zu wechseln, bestand von Anfang an das Konzept der „Dreh- / Drück- Bedienung“. Diese mechanische Bedienung beschränkte sich auf den Einsatz der Haptik. So hat der Benutzer die Möglichkeit mit einem Drehrad die Lautstärke bzw. Frequenz einzustellen und über einige Drückknöpfe direkt zu Radiosendern oder anderen Frequenzbändern zu wechseln. Dieses klassische Schnittstellendesign wurde im Laufe der Zeit verfeinert und so sind auch die meisten Navigationssysteme heute über Knöpfe und Drehräder zu bedienen. Hierbei wurde aber zum Beispiel für die Eingabe von Stadtnamen das Konzept erweitert. Der so genannte *Automatic Speller* ermöglicht per Drehrad im Alphabet einen Buchstaben einzustellen, diesen dann per Druck auf das Rad auszuwählen und blendet im Folgenden alle zur Verfügung stehenden Buchstaben für die folgende Stelle ein. So muss der Benutzer im weiteren Verlauf nur noch aus den möglichen Folgebuchstaben wählen und verkürzt so die Eingabezeit. Ein Beispiel für ein Radionavigationsgerät mit solch einer Bedienmodalität ist der TravelPilot Rome der Blaupunkt GmbH (s. Abb. 2.2)



Abbildung 2.2: Travel Pilot Rome. Bedienung im Dreh-/Drück-Konzept

Parallel zu der Verfeinerung dieses klassischen Bedienkonzepts erhalten auch neue Technologien Einzug in die Mittelkonsole. So sind gerade mobile Navigationssysteme, aber auch Erstaustattungsgeräte der höheren Qualität häufig per Touchscreen zu bedienen. Mit Bedienstift oder dem Finger kann der Fahrer so jegliche Funktionen des Geräts auswählen. Bei der Eingabe von Text wird wiederum auch auf Bestehendes zurückgegriffen. So muss der Benutzer jeden Buchstaben aus einer angezeigten Tastatur auswählen. Zwar kommt hierbei teilweise auch das intelligente Ausblenden nicht verfügbarer Buchstaben zum Einsatz, der Aufwand jeden Buchstaben einzeln einzugeben, besteht aber weiterhin.

Sind die Benutzerschnittstellen dieser Einzelgeräte noch so einfach zu bedienen, besteht dennoch das Problem, dass der Fahrer mehrere Geräte in der Mittelkonsole hat, die er teilweise während der Fahrt bedient. Der damit verbundene Aufwand und die Ablenkung von der Primäraufgabe sind enorm. Wie eingangs des Kapitels erwähnt, sind also Automobilhersteller und –zulieferer bemüht, schon bei der Erstausrüstung eines Fahrzeugs, alle Informations- und Kommunikationsfunktionen in ein System zu integrieren. Über ein solches Fahrerinformationssystem hat der Kunde dann Zugriff auf Telefon, Radio, CD-Wechsler, Navigationssystem und eventuell weitere Funktionen wie Internetbrowsing oder Email. Die oben vorgestellte Bediensystematik stößt bei solch einer Vielzahl von Optionen schnell an ihre Grenzen. Zwar gibt es auch auf diesem Gebiet Systeme, die allein per Dreh und Druck bedient werden, aber dennoch führt der Anspruch des Benutzers, ein FIS so schnell und einfach wie möglich bedienen zu können, die Entwickler solcher Systeme zu einer anderen Bedienmodalität: der Sprache.

Die Sprachbedienung bietet die Möglichkeit, Aufwand und somit Ablenkung zu minimieren. Der einzige manuelle Knopfdruck des Fahrers findet bei heutigen Sprachbediensystemen am Lenkrad statt. Der so genannte Push-To-Talk (PTT) Knopf aktiviert das Mikrofon und ermöglicht dem Nutzer per Sprachbefehl auf alle Funktionalitäten des FIS zuzugreifen. Das System kann hierbei unterstützen, indem es die zur Verfügung stehenden Sprachkommandos auf dem Display anzeigt oder nach Aufforderung auch vorliest. Durch das Vorlesen der Befehle und aktiver Rückfrage bei nicht oder falsch verstandenen Eingaben entsteht zunehmend ein

Dialog zwischen System und Nutzer. Wie aber solch ein Sprachdialog konzipiert sein muss, um die oben erklärten Maxime der effektiven und effizienten Bedienbarkeit sowie der Nutzerakzeptanz zu erfüllen, muss in einer Erweiterung der klassischen Usability-Richtlinien definiert werden. (s. *Abschnitt 2.3.3.*) Die Verwendung einer dem Nutzer vertrauten Modalität lässt zumindest vermuten, dass es bei angemessenem Einsatz eine schnelle Erlernbarkeit und eine hohe Akzeptanz ermöglicht. Diese Überlegung führt auch zu weiteren Eingabemöglichkeiten, wie z.B. der Gestik oder Mimik. Auch hier lassen sich alltägliche Bedeutungen für den Gebrauch im Fahrzeug ableiten und helfen dem Benutzer, mit möglichst wenig Aufwand auf eine Komplexität an Funktionen zuzugreifen.

### **2.3.2. Multimodale Fahrerinformationssysteme**

Festzustellen ist, dass neben der mechanischen Bedienung von Fahrerinformationssystemen weitere Modalitäten wie Sprache, Gestik oder auch Mimik viel versprechende Möglichkeiten bieten, „die Bedienung komplexer Systeme intuitiver [zu] gestalten. Ebenso ist eine höhere Bedienungseffizienz zu erwarten, weil unabhängige Ressourcen zum Einsatz kommen.“ (AKYOL et al. 2001: 138) Dem großen Umfang an verfügbaren Funktionen kann so eine ebenso große Vielfalt an Bedienmodalitäten gegenüber gestellt werden. Dies bedeutet für den Nutzer keine Komplexitätssteigerung in der Bedienung, sondern ermöglicht es ihm, für die jeweilige Funktion eine angemessene Steuerungsmöglichkeit zu haben. Gerade die Kombination der verschiedenen Eingabemechanismen in einer Benutzerschnittstelle verspricht eine große Usability des Systems. Gibbon et al. stellen hierzu folgende Vorteile multimodaler Schnittstellen fest:

- *Modality synergy*

Eine Nutzereingabe über mehrere Modalitäten ermöglicht eine genauere Interpretation der Nutzerintention seitens des Systems.

- *Different modalities, different benefits*

Der Zugriff über mehrere Modalitäten erlaubt es die für die jeweilige Funktion passende Eingabe zu verwenden. So ist bei der Eingabe von Namen die Sprache besser geeignet, bei der Auswahl einer Option eher die Deiktik.

- *New applications*

Einige Funktionen werden durch die Hinzunahme von Modalitäten erst möglich. So erlaubt erst die Sprachein- und -ausgabe das Diktieren bzw. Vorlesen von Emails.

- *Freedom of choice*

Auch wenn dieselbe Aufgabe mit gleicher Effizienz durch die eine oder andere Modalität zu erledigen ist, so haben Nutzer doch spezifische Präferenzen bei der Modalitätenwahl. Multimodale Schnittstellen erlauben es jedem Nutzer, die von ihm bevorzugte Variante zu wählen.

- *Naturalness*

Das Angebot multipler Modalitäten zur Interaktion mit einem System kann natürlicher für den Menschen sein, wenn Gewohnheiten und Strategien der Mensch-Mensch Kommunikation auf die Mensch-Maschine-Interaktion übertragen werden.

- *Adaptation of either several possible environmental settings or evolving environments*

Die Möglichkeit von einer Modalität zur anderen je nach äußeren Bedingungen (Lärm, Licht,...) zu wechseln, ist ein weiterer Vorteil multimodaler Systeme.

(vgl. 2000: 106ff.)

Gerade der letzte Aspekt macht deutlich, dass im automobilen Umfeld ein bedeutendes Potenzial in der Erweiterung bisheriger Bedienmodalitäten liegt. Durch ständig wechselnde situative Beanspruchung des Fahrers innerhalb seiner Primäraufgabe ist die Verfügbarkeit mehrerer Eingabekanäle von großer Bedeutung. *Abbildung 2.3* zeigt wie ein solches Bedienkonzept aufgebaut sein könnte und welche Ein- und Ausgaben darin enthalten sind.

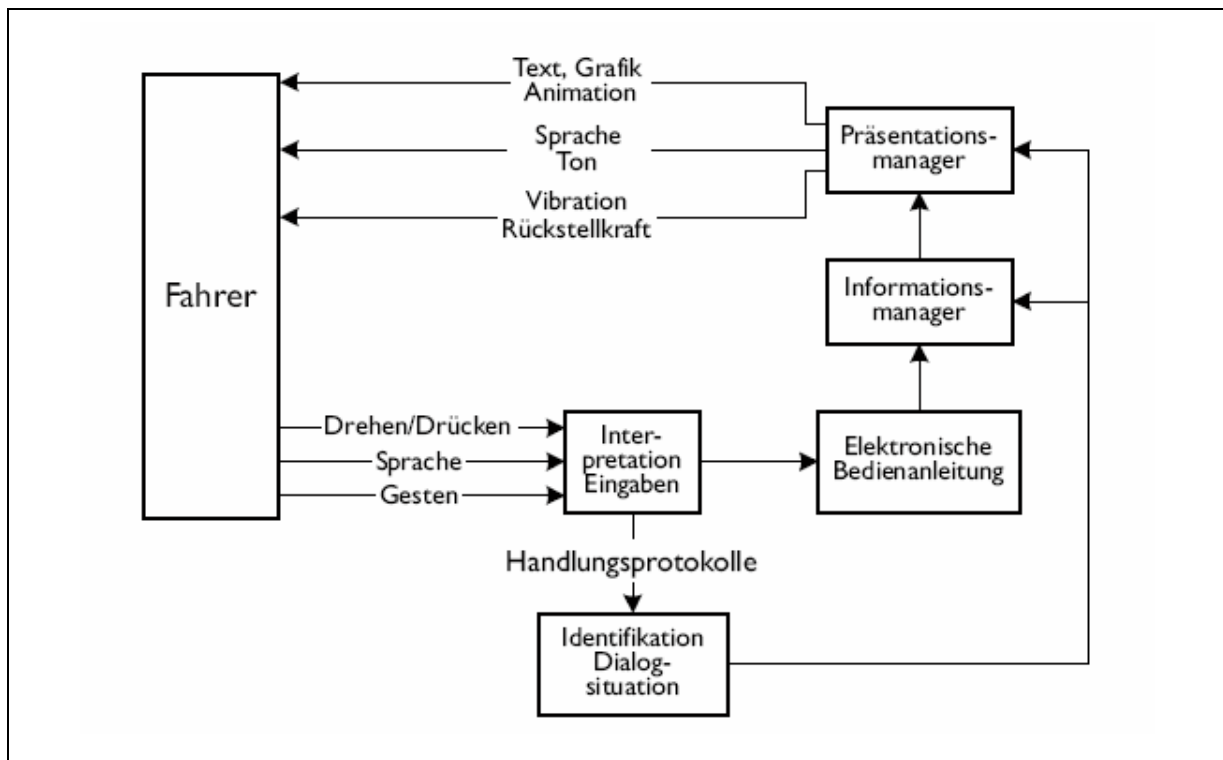


Abbildung 2.3: Konzept eines multimodalen Kfz-Informationssystems

(AKYOL et al. 2001: 144)

Wie in der Abbildung ersichtlich, sieht das Konzept der Multimodalität aber neben mehreren Eingabekanälen (Drehen / Drücken, Sprache, Gesten) auch entsprechende Ausgabeformen vor. Die dadurch entstehende Symmetrie der Benutzungsschnittstelle soll sicherstellen, dass nicht nur die Steuerung sondern auch die Informationsaufnahme erleichtert werden kann.

Für die vorliegende Untersuchung sind vor allem zwei Aspekte dieser Grafik von Bedeutung. Dies ist zum einen die Technologie der Sprachein- und -ausgabe. Es scheint sich gerade für die Eingabe von Navigationszielen in der Routenführung die gesprochene Sprache als geeignete Modalität anzubieten. Auch in anderen Situationen, in denen der Fahrer seine Hände nicht vom Lenkrad nehmen kann oder möchte, ist eine Sprachbedienung die optimale Alternative. Zusätzlich wird in diesem Schema aber auch die Möglichkeit der animierten Ausgabe in Betracht gezogen. Inwiefern gerade dies ein potenzielles Werkzeug zur Gestaltung eines Systems von hoher Usability sein könnte, wird näher in *Kapitel 3* erläutert.

Der Einsatz von Sprache im Fahrzeug zur Bedienung von Fahrerinformationssystemen scheint also unter Anbetracht der Anwendungssituation ein entscheidendes Bedienelement zu sein. Wie aber solch eine Modalität eingesetzt werden kann, um auch den Anspruch zu erfüllen, die Usability für den Benutzer zu erhöhen, gilt es näher zu definieren. Die einleitenden internationalen Richtlinien sind hierfür zu allgemein. Der folgende Abschnitt stellt einen erweiterten Begriff der Usability in Bezug auf Sprachbedienung vor.

### **2.3.3. Usability und Akzeptanz von Sprachdialogsystemen**

Durch den Einsatz eines multimodalen Fahrerinformationssystems bietet sich die Möglichkeit, die angesprochenen Geräte und Funktionalitäten dem Benutzer mit den oben genannten Vorteilen der Bedienung zugänglich zu machen. Allerdings gelten gerade beim Einsatz von Sprache spezifische Richtlinien. Für die Gestaltung eines interaktiven Sprachdialogsystems gilt es also, die in *Kapitel 2.1.* definierten Grundsätze der Usability zu erweitern.

Im Allgemeinen muss ein Sprachdialogsystem Benutzeranforderungen ähnlich denen gewöhnlicher interaktiver Systeme erfüllen. Jedoch gibt es grundlegende Unterschiede, die auch ihre Berücksichtigung bei der Konzeption solcher Systeme finden müssen. So ist z.B. die gesprochene Sprache in der Wahrnehmung ein flüchtiges Phänomen und nicht statisch. Der Benutzer muss also mit erhöhter Aufmerksamkeit die Informationsausgabe des Systems beachten. Es bedeutet auch, dass er das Dialogsystem im Vorfeld nicht grundlegend inspizieren und sich damit vertraut machen kann. Ist die Interaktion mit einem solchen System nicht selbsterklärend konzipiert, so muss der Benutzer nach der Trial-and-Error-Methode<sup>4</sup> vorgehen und nach jedem fehlgeschlagenen Bedienversuch wieder von vorne beginnen. Des Weiteren muss auch bei der Spracheingabe berücksichtigt werden, dass die Sprachverarbeitung und das Sprachverstehen schwieriger und fehleranfälliger sind als die Interpretation haptischer Eingabe. So sollte im besten Falle die Gestaltung eines Sprachdialogsystems bereits so konzipiert sein, dass

---

<sup>4</sup> Methode, den besten Weg zur Lösung eines Problems zu finden, indem verschiedene Möglichkeiten ausprobiert werden



vermieden werden kann, dass der Benutzer fehlerhafte oder missverständliche Eingaben produziert. (vgl. DYBKJAER / BERNSEN 2000: 245) In einer Studie zur Akzeptanz und Usability von Sprachapplikationen in Deutschland entwickeln Peissner et al. sechs Schlussfolgerungen für eine Erweiterung der klassischen Usability-Richtlinien:

- *„Kurz und Einfach*

Reduzieren Sie die Länge der ausgegebenen Texte auf das notwendige Minimum: Kurze Sätze, wenige, kurze Items in einer Liste, einfacher Satzbau, keine umständlichen Formulierungen

- *Der Benutzer kontrolliert den Dialog*

Stellen Sie sicher, dass der Benutzer sich stets als aktiver Gesprächsführer empfindet. Barge-In und andere Steuerungsfunktionen wie „Weiter“ und „Zurück“ sind dabei unerlässlich. Neue und gelegentliche Benutzer lassen sich gerne durch Menüs „führen“. Bieten Sie jedoch dem Benutzer stets die Möglichkeit, selbst die Initiative zu ergreifen (Mixed Initiative) und ermöglichen Sie „Abkürzungen“ durch die Menüstruktur.

- *Die Sprache des Benutzers*

Sprechen Sie die Sprache des Benutzers. Vermeiden Sie fremde Begriffe und unbekannte Markennamen. Stellen Sie sicher, dass der Benutzer mit einfachen und intuitiven Äußerungen an sein Ziel gelangt.

- *Design for Error*

Berücksichtigen Sie die Möglichkeit von Erkennungsfehlern in allen Aspekten der Dialoggestaltung. Der Benutzer soll stets Erkennungsfehler bemerken und sie schnell und einfach korrigieren können.

- *Orientierung*

Stellen Sie sicher, dass der Benutzer stets weiß, wo er sich befindet, warum er hier ist, welche Optionen ihm zur Verfügung stehen und wohin er gehen kann.

- *Ästhetische Konsistenz*

Die Stimme, der Sprachstil, die anvisierten Zielgruppen, die Anwendungsdomäne, das repräsentierte Unternehmen, die Ziele und Inhalte der Applikation – das alles muss zusammen passen und ein in sich stimmiges Gesamtbild ergeben, um Kundenvertrauen zu schaffen.“ (2004: 63)

Diese Richtlinien für die Gestaltung von Sprachdialogsystemen zeigen die vielen zusätzlichen Aspekte auf, die bei der Gestaltung und Prüfung der Usability solcher Systeme zusätzlich berücksichtigt werden müssen. In der oben genannten Studie

wird jedoch deutlich, dass dies bei den bisher angebotenen Sprachapplikationen noch nicht der Fall ist. „Mehr als jeder Vierte (76 bzw. 26%) der 289 Nutzer gibt an, „sehr unzufrieden“ mit der von ihm am häufigsten genutzten Sprachapplikation zu sein.“ (a.a.O.: 17) Die logische Schlussfolgerung ist zum einen, dass die oben genannten Hinweise noch nicht ausreichend in die Gestaltung bisheriger Systeme eingeflossen sind. Zum anderen stellt sich die Frage, ob neben einer effektiven und effizienten Systemgestaltung, wie sie in den Normen und Richtlinien definiert wird, andere Eigenschaften die Nutzerakzeptanz generell erhöhen können. Peissner et al. erwähnen in diesem Zusammenhang die Bedeutung von Anmutung und emotionaler Wirkung einer Benutzerschnittstelle. „Hierfür spielen nicht nur die oben genannten ergonomischen Aspekte eine große Rolle, sondern ganz besonders auch die Merkmale des Systems, die den Kunden emotional ansprechen und damit spontan zu einer ablehnenden oder positiven Grundhaltung führen.“ (a.a.O.: 61) In der Usability-Forschung gewinnt zunehmend das dritte Charakteristikum des Usability-Begriffs an Bedeutung: Die Akzeptanz. Hassenzahl und Hofvenschiöld formulieren ein wenig provokant: „If it doesn't feel right, who cares if it works? Oder muss Software mehr als nur gebrauchstauglich sein?“ (2003: 135) Diese Perspektive setzt einen stärkeren Fokus auf Anmutung, Design und Spaß als weitere Kriterien für Usability. In einer Erweiterung der bestehenden ISO Norm 9241 für Multimedia-Benutzerschnittstellen, der ISO Norm 14915 Teil 1, finden sich Hinweise für die Bedeutung dieser Überlegung. „Falls es für die Arbeitsaufgabe angebracht ist, sollte eine Multimedia-Anwendung so gestaltet sein, dass sie für den Benutzer anregend ist, d.h. dass sie die Aufmerksamkeit des Benutzers auf sich zieht und ihn dazu motiviert, mit ihr zu interagieren.“ (DIN 2004: 336)

Wie dieses Kapitel aufgezeigt hat, sind die klassischen Prinzipien einer benutzergerechten Systemgestaltung besonders in Bezug auf den Kontext und die Anwendung zu konkretisieren. Für Sprachdialogsysteme heißt das, dass die zusätzlichen Technologien der Spracherkennung und Sprachausgabe spezifische Richtlinien erfüllen müssen, aber auch die Dialogstruktur und die Anmutung einer Anwendung dazu beitragen, dass der Benutzer gerne, effektiv und effizient damit interagiert.

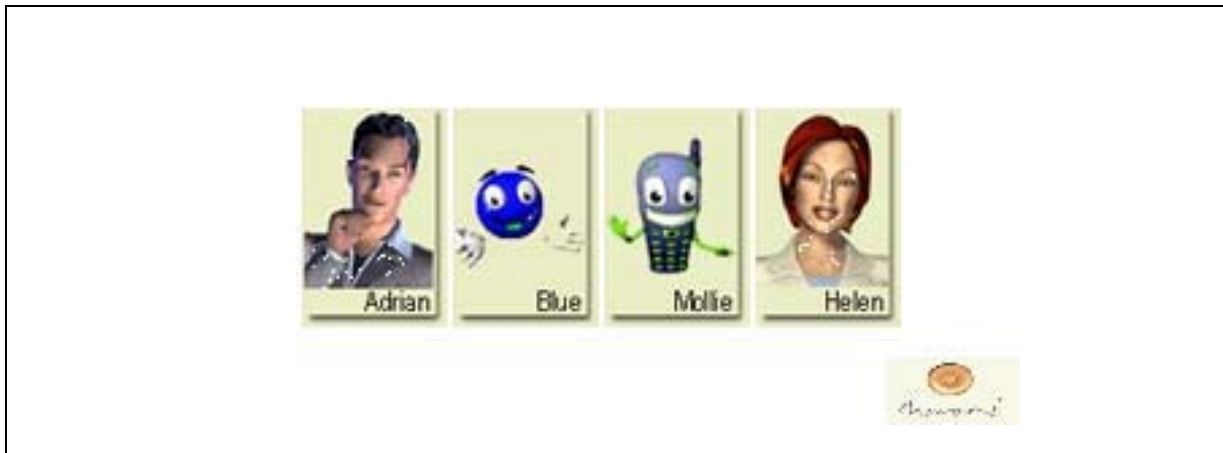
### **3. Anthropomorphe Interface-Agenten (AIA)**

Ein anthropomorpher Interface-Agent verspricht ein mögliches Mittel zu sein, um die in *Kapitel 2* beschriebenen Anforderungen an die Gestaltung von sprachbedienbaren Benutzerschnittstellen zu erfüllen. Was genau ist aber mit diesem Begriff gemeint? Wie kann diese Technologie als Benutzerschnittstelle eingesetzt werden? Und inwiefern erhofft man sich gerade für ein sprachbedienbares Fahrerinformationssystem dadurch eine Steigerung der Nutzerakzeptanz oder gar der Usability? Das folgende Kapitel klärt im ersten Teil Begrifflichkeiten um diese Technologie und zeigt auf, was solche Agenten sind und wo sie eingesetzt werden können. In *Abschnitt 3.3* werden Beweggründe zum Einsatz in einem Sprachdialogsystem erläutert und Hypothesen bzw. Fragestellungen für die Untersuchung aufgestellt.

#### **3.1. Begriffsabgrenzung**

Bei der hier vorgestellten Technologie handelt es sich um animierte Charaktere. „Diese menschenähnlich gestalteten und autonom interagierenden virtuellen Figuren sollen im Rahmen der Mensch-Technik-Interaktion zukünftig den Umgang mit beispielsweise Computer oder Videorekorder erleichtern“. (BLENS et al. 2003: 297) In dieser Definition findet man grundlegende Merkmale, die anthropomorphe Interface-Agenten von anderen Begriffen wie zum Beispiel einem *Avatar* unterscheiden. Avatare sind eher definiert als „Stellvertreter, virtuelle Repräsentanten der materiegebundenen Wesen, die sie ins Leben gerufen haben“. (DÖRING 1999: 98) Sie waren die ersten virtuellen Charaktere und tauchten vornehmlich in Chaträumen, Multi-User-Dungeons, Online Spielen oder Virtual Reality Umgebungen auf. Avatare basieren zwar auf derselben Technologie, nämlich einer heute meist 3D-generierten Animation, erfüllen jedoch einen anderen Zweck, nämlich den der Repräsentation eines Chat-Users oder eines Online-Spielers. Anthropomorphe Interface-Agenten (AIA) können dem englischsprachigen Begriff *Embodied Conversational Agents* (ECA) gleichgestellt werden. (vgl. KRÄMER / BENTE 2002:

205ff.) Diese Bezeichnung macht den Unterschied zu reinen Animationen zur Unterhaltung deutlich. Er liegt vor allem im Zweck und dem Einsatzbereich: „an embodied conversational agent [is] an interface in which the system is represented as a person, information is conveyed to human users by multiple modalities such as voice and hand gestures“. (CASSELL 2001: 67)



**Abbildung 3.1: Virtuelle Charaktere der Charamel GmbH**

Die Eigenschaft *anthropomorph* bzw. *embodied* meint also, dass dieser Agent *menschenähnlich* ist bzw. eine *Verkörperung* erhält. Dabei muss er nicht unbedingt die Abbildung einer speziellen Person sein, sondern hat manchmal lediglich charakteristische Körpereigenschaften, die dem der Menschen nachempfunden sind. Die Charamel GmbH, die auch den Prototypen dieser Studie zur Verfügung gestellt hat, bietet ihren Kunden eine Vielfalt solcher Charaktere an. *Abbildung 3.1* zeigt beispielhaft einige ihrer Figuren, wobei deutlich wird, dass gerade die menschlichen Körperteile, die die zwischenmenschliche Kommunikation unterstützen, zentrale Elemente solcher Visualisierungen sind. So werden auch Gegenstände wie der Punkt oder das Mobiltelefon mit den Elementen der nonverbalen Kommunikation ausgestattet, um die menschliche Mimik und Gestik nachzuahmen.

### **3.2. Bisherige Einsatzbereiche virtueller Charaktere**

Durch die oben vorgenommene Begriffsabgrenzung ist deutlich geworden, dass ein AIA eine zentrale Funktion hat. Er dient zur Gestaltung von Sprachdialogen oder zumindest zur Informationsrepräsentation in Kombination mit gesprochener Sprache. Diese Hauptaufgabe lässt sich in verschiedenen Einsatzbereichen beobachten. Einer der bekanntesten Charaktere innerhalb der Informationswissenschaft ist *REA – the real estate agent*. Dieses Projekt der *Gesture and Narrative Language Group* des *MIT Media Laboratory* entwickelte eine digitale Immobilienmaklerin, um multimodale Mensch-Maschine-Interaktion im Kontext eines Verkaufsgesprächs so nah wie möglich der zwischenmenschlichen Kommunikation nachzuempfinden. (vgl. hierzu CASSELL 2001) Diese wissenschaftliche Konzeption zeigt einen Anwendungsbereich auf, der auch kommerziell bereits ansatzweise erschlossen ist. Die Charamel GmbH setzt zum Beispiel ihre Charaktere bereits als Verkaufsassistenten ein. An so genannten *Points-of-Sale* (POS) kann sich der Kunde durch die Interaktion mit einem AIA über Produkte und Angebote eines Unternehmens informieren. Dabei setzen die Auftraggeber solcher Informationsstationen auf die hohe Kundenattraktivität und – aufmerksamkeit, die eine virtuelle Verkäuferin erzielt. So führte der Einsatz solcher Systeme zur Produktmoderation in einer Kaufhauskette zu einem Anstieg der Kundenansprachen um das 30-fache und eine Erhöhung der Verweildauer. (vgl. STRICKER 2003: 179) Der virtuelle Charakter weckte also schon im Vorfeld die nötige Neugier und Attraktion, um die Benutzung des Systems zu motivieren. Im Rahmen der Interaktion stellen dann die persönliche Ansprache, individualisierte Informationen und die Anmutung des Charakters einen emotionalen Mehrwert dar. Diese Art von „Erlebnis-Marketing“ ist ein heute schon erfolgreicher Einsatzbereich von AIA.

Auch auf Internetseiten finden sich teilweise bereits animierte *Guides*, die durch den Inhalt der Seiten führen und als Ansprechpartner dem User zur Verfügung stehen. Der Einsatz aufwändiger Animationen auf Webseiten entwickelte sich parallel zu der verfügbaren Bandbreite, und bei den heutigen Datenübertragungsraten lösen sich gerade ehemalige technologische Hemmnisse auf und machen den Weg für virtuelle Charaktere frei. Bereits 2001 ergab eine Studie des NFO Infratest und dem

Fachverlag Symposium Publishing: „40 Prozent der deutschen Internet-User finden eine interaktive Produktberatung durch virtuelle Charaktere wichtig oder sehr wichtig“. (a.a.O.: 171) Auch hier ist der empfundene Mehrwert keineswegs in Verbindung mit Usability zu bringen, sondern eher eine Frage der Kundenattraktivität. Dieser Aspekt ist jedoch nicht zu vernachlässigen, denn gerade ein personalisierter, animierter Vertreter eines Unternehmens kann cross-medial als Werbeträger eingesetzt werden und für ein positives Firmenimage sorgen. Die heute grafisch oft hochwertige Darstellung hat also einen hohen Werbeeffect. Das ansprechende Design in Zusammenklang mit unterhaltsamen Inhalten eröffnet auch neue Märkte, wie zum Beispiel den der mobilen Anwendungen. Videonachrichten per Mobiltelefon, vorgetragen durch einen virtuellen Charakter sind dabei nur eine mögliche Anwendung innerhalb der Unterhaltungsindustrie.

Im Vordergrund bisheriger Einsatzbereiche steht die Intention, Kundenaufmerksamkeit zu gewinnen und in simulierten Verkaufsgesprächen die Vorteile personifizierter Informationssysteme zu nutzen, um ein Produkt zu bewerben und den Absatz zu steigern. Es festigt sich der Eindruck, dass sich auf diesem Gebiet der Einsatz virtueller Charaktere als sehr erfolgreich herausstellt. Es stellt sich die Frage, wie sich die bisher positiven Erkenntnisse bezüglich Kundenakzeptanz –und aufmerksamkeit auf andere Einsatzbereiche übertragen lassen. Die möglichen Potenziale des Einsatzes AIA im Sprachdialog eines Fahrerinformationssystems sollen im nächsten Kapitel vorgestellt werden. Diese Überlegungen münden in einige Hypothesen, die es in der Evaluierung der Prototypen zu prüfen gilt.

### **3.3. Potenziale für den Sprachdialog im Fahrerinformationssystem**

*“As Don Nielson [...] used to say, talking to a computer without an animated persona who can speak back is like talking to a wall.” (CHEYER / JULIA 1999: 1)*

Die Idee, anthropomorphe Interface-Agenten in einem sprachbedienbaren Fahrerinformationssystem einzusetzen, beruht auf zwei Überlegungen. Da ist zum einen die Frage nach der benutzergerechten Gestaltung von Sprachdialogsystemen.

*Kapitel 2.3.3.* hat zwar einerseits Richtlinien für die Gestaltung von Sprachdialogsystemen aufgezeigt, aber auch darauf hingewiesen, dass die Nutzerzufriedenheit sprachbedienbarer Systeme häufig noch ausbleibt. Die Skepsis der Nutzer ist vermutlich auf Erfahrungen mit unzureichend gestalteten und konzipierten Dialogsystemen zurückzuführen. Ein nicht zu vernachlässigender Aspekt wird aber auch im obigen Zitat verdeutlicht. Mit der Weiterentwicklung der Spracherkennungs- und ausgabetechnologie beginnen sich Computersysteme vom maschinellen und unnatürlichen hin zu einem flüssigen und natürlichsprachlichen Dialog zu entwickeln. Die bisherige sprachliche Interaktion mit einem Computer war davon geprägt, dass die Anwender in der Bedienung eines in der Kommunikation eher unnatürlichen Systems ihr Eingabeverhalten stark änderten und der Maschine anpassten.<sup>5</sup> Heutige Sprachdialogsysteme erzeugen eine Mensch-Maschine-Interaktion, die zunehmend der zwischenmenschlichen Kommunikation nahe kommt. Erlaubt also die technische Qualität eines Dialogsystems sprachliche Eingaben in Form längerer Befehl –oder Satzkonstrukte, dann befindet sich der Nutzer tatsächlich in einer Gesprächssituation. Eine Darstellung der Anwendung durch Fenster, Listen und Symbole bietet dem Benutzer jedoch nicht die für eine solche Situation gewohnte nonverbale Rückmeldung in Form von Gesten, Mimik oder Körpersprache eines Gegenübers, die elementar für eine erfolgreiche Gesprächsführung sind. So fällt es schwer, die Situation als solche zu erfassen und die Interaktion darauf einzustellen. Der Nutzer spricht mit einer *schwarzen Kiste*. Daraus folgen verfrühte oder falsche Eingaben und eine allgemeine Unsicherheit im Umgang mit dem System. Befürworter anthropomorpher Schnittstellen fordern demnach ein menschliches Antlitz für solche Systeme. CASSEL et al. formulierten im Rahmen des oben erwähnten Projekts REA: „*We believe that interfaces that are truly conversational have the promise of being more intuitive to learn, more resistant to communication breakdown, and more functional in high ‚noise‘ environments. Therefore, we propose to leverage the full breadth and power of human conversational competency by imbuing the computer with all of the conversational skills that humans have; to wit, the ability to use the face, hands, and melody of voice to regulate the process of conversation*” (1999: 520) Der Einsatz eines AIA in einem sprachbedienbaren Fahrerinformationssystem wäre

---

<sup>5</sup> (vgl. hierzu *Computer Talk Hypothese*: KRAUSE et al 1992)

diesen Stimmen zufolge nur die logische Konsequenz aus gesteigerter Sprachdialog-Qualität und der Intention, dem Fahrer die Bedienung des Systems so intuitiv wie möglich zu gestalten. Der Einsatz nonverbaler Elemente der zwischenmenschlichen Kommunikation ermögliche die Simulation realen Gesprächsverhaltens und erziele so eine fehlerfreie Bedienbarkeit. Dieser erste Aspekt betrifft also eine mögliche Verbesserung der Usability des Sprachdialogsystems durch die Verwendung virtueller Charaktere.

Der zweite Beweggrund für die prototypische Umsetzung eines virtuellen Informationsagenten im Fahrzeug sind die Ergebnisse bisheriger Einsatzbereiche. Das offensichtliche Interesse an solchen Charakteren und das hohe Marketingpotenzial dieser Technologie können zu einem Differenzierungsmerkmal im Markt der Fahrerinformationssysteme der gehobenen Preisklasse führen. Zusätzlich lässt sich aber auch vermuten, dass sich gerade die hohen Kundenakzeptanzwerte bei Verkaufssystemen ebenfalls positiv auf die Verwendung in Sprachdialogsystemen niederschlagen. Erste Studien zeigten, „dass die Möglichkeit der Spracheingabe signifikant häufiger genutzt wird, wenn ein menschenähnliches Gesicht auf dem Bildschirm dargeboten wird.“ (KRÄMER / NITSCHKE 2002: 231) Der in *Abschnitt 3.2* erwähnte *emotionale Mehrwert* könnte sich demnach alleine schon auf Akzeptanz und Verwendungsbereitschaft der Anwender positiv auswirken. Eine ansprechend animierte Bedienoberfläche eines Fahrerinformationssystems kann zusätzlich bewirken, dass der Fahrer mehr Freude an der Benutzung eines Sprachdialogs empfindet. So entspricht der Einsatz virtueller Charaktere auch den erweiterten Usability-Richtlinien, die einen emotionalen Anspruch und Nutzer-Motivation bzw. „Joy of Use“ fordern.

Fasst man die Erkenntnisse der bisher aufgeführten wissenschaftlichen und marktwirtschaftlichen Studien zusammen, so lassen sich für den Einsatz eines anthropomorphen Interface-Agenten in einem multimodalen Fahrerinformationssystem folgende allgemeine Hypothesen formulieren:



- *Hypothese H1*

Die Verwendung eines menschenähnlichen Charakters im Sprachdialog eines Fahrerinformationssystems simuliert eine Situation zwischenmenschlicher Kommunikation und bewirkt somit eine intuitivere Bedienbarkeit für Erstnutzer.

- *Hypothese H2*

Die qualitativ hochwertige und aufwändige Animation eines AIA steigert das Kundeninteresse und somit die Akzeptanz der Sprachbedienung im Fahrzeug.

- *Hypothese H3*

Die Bedienung des FIS im Dialog mit einem virtuellen Charakter bereitet dem Nutzer mehr Freude als mit einer Benutzeroberfläche ohne Animation. Die so gesteigerte Motivation und Entspannung des Anwenders bewirkt weniger Bedienfehler und eine größere Zufriedenheit mit dem System.

## 4. Usability-Evaluierung eines AIA im multimodalen FIS

„Walker et al. (1994) beschreiben die hier gestellte Aufgabe treffend: „The goal of HCI work with synthetic faces should not necessarily be to give a computer a human face but rather to determine when a face-like interface is appropriate“ (S. 90)<sup>6</sup>. Zur Beantwortung dieser Fragen empfiehlt sich ein experimentelles Vorgehen: Als abhängige Variablen bzw. Evaluationskriterien lassen sich Effektivität (Grad der Zielerreichung), Effizienz (Grad der Zielerreichung in Bezug zu aufgewendeten Ressourcen) sowie Akzeptanz (subjektive Reaktion des Benutzers im Sinne von Zufriedenstellung) unterscheiden.“ (KRÄMER/BENTE 2002: 217) Eine analytische und empirische Untersuchung kann so dazu beitragen, die vorangestellten Überlegungen zum Einsatz virtueller Charaktere im Fahrzeug mit qualitativen und quantitativen Daten zu prüfen. So ist das zentrale Element dieser Arbeit die Konzeption und Durchführung einer vergleichenden Evaluierung der beschriebenen Prototypen. Der direkte Vergleich eines sprachbedienbaren Fahrerinformationssystems mit und einem ohne einen anthropomorphen Interface-Agenten ermöglicht die Analyse der unmittelbaren Auswirkungen auf die Usability und die Akzeptanz von Systemen mit solchen Agenten und somit die Prüfung der oben formulierten Hypothesen. Der erste Teil dieses Kapitels stellt zunächst die zu untersuchenden Prototypen vor. Anschließend folgt dann die Beschreibung der zweistufigen Durchführung der Evaluierung. In den beiden Abschnitten 4.3 und 4.4 werden jeweils erst der Aufbau und das Vorgehen der Experten-Evaluierung bzw. des Benutzertests erläutert, bevor dann die Ergebnisse der einzelnen Untersuchungsschritte dargelegt werden.

---

<sup>6</sup> KRÄMER und BENTE beziehen sich hier auf: WALKER, J.H.; SPROULL, L.; SUBRAMANI, R. (1994): *Using a Human Face in an Interface*. In ADELSON, B.; DUMAIS, S.; OLSON, J. (Hrsg.) (1994): *Human Factors in Computing Systems: Chi '94 Conference Proceedings*. Boston: ACM. (S. 85-91)

## 4.1. Beschreibung der Prototypen

Um die oben genannten Hypothesen zu prüfen, wurden zwei unterschiedliche Prototypen entwickelt. Diese Systeme sind PC-basierte Anwendungen, die auf einem Computer mit Mikrofon lauffähig sind. Bevor die genaue Konzeption der Untersuchung erläutert wird, dient dieser Abschnitt der Darstellung des Aufbaus und der Funktionsweise der Prototypen. Die Beschreibung der Systeme bezieht sich auf den Zustand zu Beginn der Untersuchung. *Kapitel 4.3.3.* stellt im Verlauf der Arbeit einige Anpassung dieser Prototypen dar.

### 4.1.1. SDS Demonstrator

Das erste untersuchte System ist ein Prototyp des Sprachdialogsystems (SDS) der Blaupunkt GmbH. (s. *Abb. 4.1*) Der zu Testzwecken entwickelte *SDS Demonstrator* bietet eine grafische Oberfläche zur Evaluierung der Spracherkennung, der Sprachausgabe (TTS) und der Dialogstruktur. Diese VisualBasic-Oberfläche ist in zwei Fenster aufgeteilt.

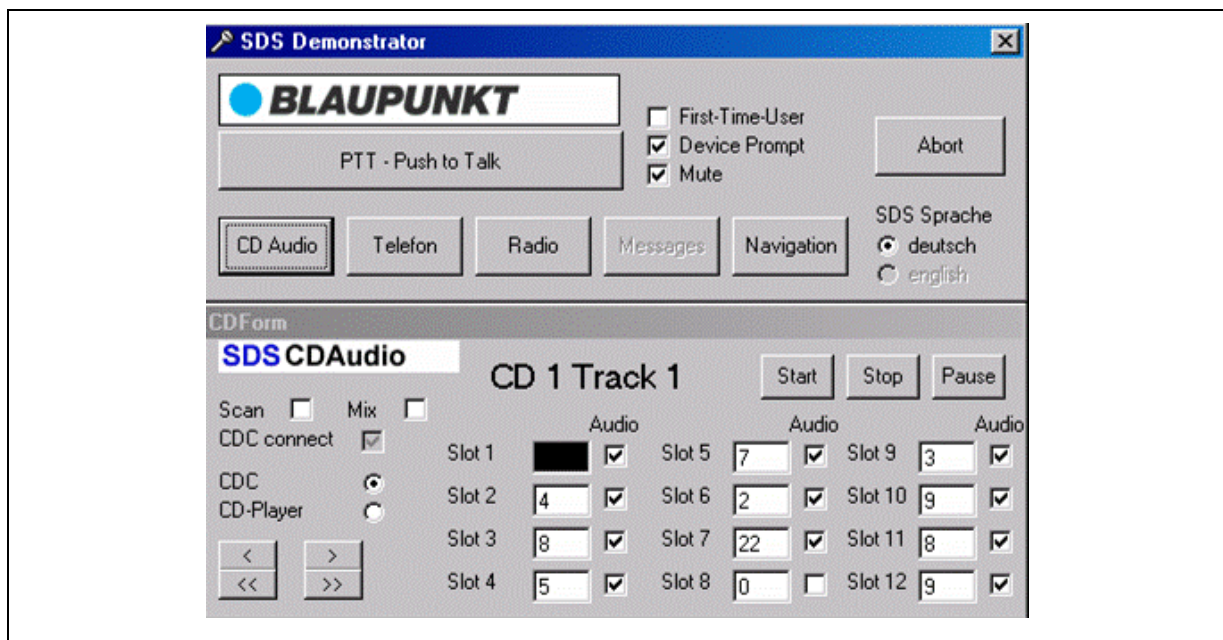
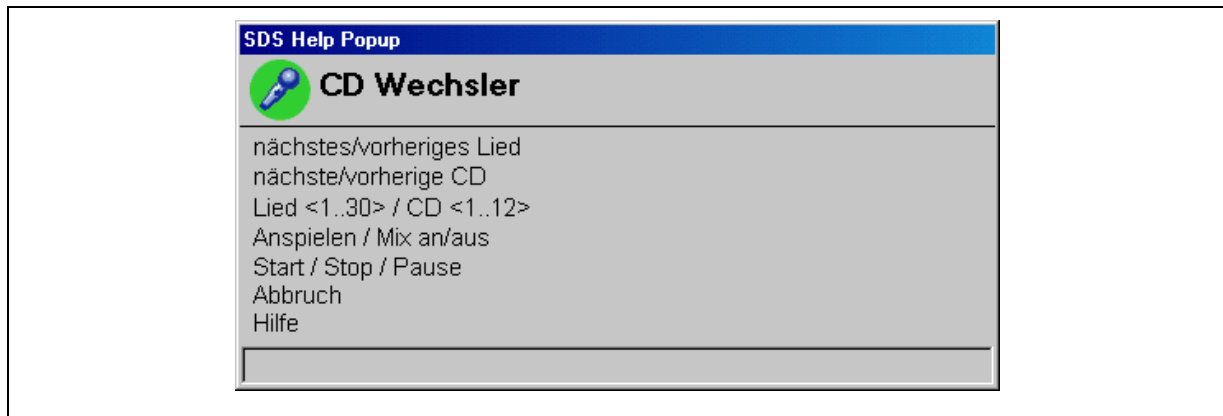


Abbildung 4.1: SDS Demonstrator

Das obere Fenster visualisiert die verschiedenen Gerätekomponten des Fahrerinformationssystems, die Systemeinstellungen und bietet zudem die Möglichkeit das Programm per Maus zu konfigurieren und zu beenden. Der *PTT – Push to Talk* - Button symbolisiert einen Druckknopf am Lenkrad, mit dem der Benutzer das System aktivieren kann. Die Checkboxes dienen der Steuerung einiger Einstellungen, wie z.B. das Ausstellen der Musikausgabe oder das Abspielen einer Einführung in das Sprachdialogsystem. Die Radio-Buttons zeigen an, welche Sprachversion gestartet wurde.

Das untere Fenster (in *Abbildung 4.1: „CD-Form“*) stellt die aktive Applikation dar. Im Falle des CD-Wechslers werden die einzelnen CD-Schächte und die Anzahl der Lieder dargestellt. Die Visualisierung ist sehr prototypisch und ist ansatzweise der Oberfläche eines CD-Spielers nachempfunden. Das Abspielen eines Liedes wird jedoch lediglich durch eine Markierung angedeutet, wohingegen bei der Radio-Applikation MP3-Dateien hinterlegt sind, die den Empfang von Radiosendern implizieren. Auch die Darstellung des Autotelefons und des Navigationssystems sollen nur grundlegendes Feedback darüber geben, was in einem wirklichen Fahrerinfoermentationssystem die jeweilige Reaktion auf Eingaben wären. Im Falle einer Aktivierung der Spracheingabe, also nach gedrücktem PTT-Knopf, wird das untere Fenster von einem Aktionsfenster (*Help Form*) überlagert. (s. *Abb. 4.2*) Diese Anzeige dient dazu, dem Anwender die zur Verfügung stehenden Befehlsphrasen anzuzeigen. Das Mikrofonsymbol und ein Signalton zeigen an, dass der Benutzer nun einen Befehl sprechen kann. Neben dem Mikrofon werden das zurzeit aktive Gerät und darunter die kontextsensitive Befehlsliste der zur Auswahl stehenden Sprachbefehle angezeigt.



**Abbildung 4.2: SDS Help Popup bei aktiver Spracheingabe**

Dieses Fenster ist daher das eigentliche Kernelement des SDS, da in einer späteren Umsetzung eines solchen Systems im Fahrzeug genau diese Informationen dem Fahrer in einem Display entweder in der Mittelkonsole oder im Kombiinstrument angezeigt würden. Ruft der Anwender eine Funktion auf, so erscheinen die für die Bedienung dieses Geräts relevanten Befehle und das Mikrofon signalisiert zeitgleich, dass das System auf eine Eingabe wartet. Bleibt dann eine Eingabe durch den Benutzer aus, so wird er wiederholt um eine Eingabe eines Sprachkommandos gebeten und im Anschluss vom System selbstständig über die zur Verfügung stehenden Kommandos informiert.

Die in diesem Prototyp eingesetzte Sprachtechnologie wird als „Sprachtechnologie Level 3“ bezeichnet und beinhaltet die Spracherkennung festgelegter Befehlsphrasen sowie die Sprachausgabe mit einem firmeneigenen Stimmenprofil auf. Die dem System hinterlegte Navigationsstruktur führt zu einer dialogischen Bedienung. Nachdem der Benutzer per Knopfdruck (PTT) eine Bediensequenz initiiert hat, wird er in aufeinander folgenden Dialogschritten durch die Bedienung des Systems geführt. Eine *Bediensequenz* ist dabei eine Aneinanderreihung von Befehlsschritten bis zum Erreichen eines Terminalbefehls. Ein solcher *Terminalbefehl* beendet die jeweilige Sequenz und führt anschließend eine konkrete Anwendung des gewählten Geräts durch. Eine mögliche Bediensequenz, z.B. zum Abspielen einer CD aus dem dritten Schacht des CD-Wechslers, sieht demnach wie folgt aus. (s. Abb. 4.3)

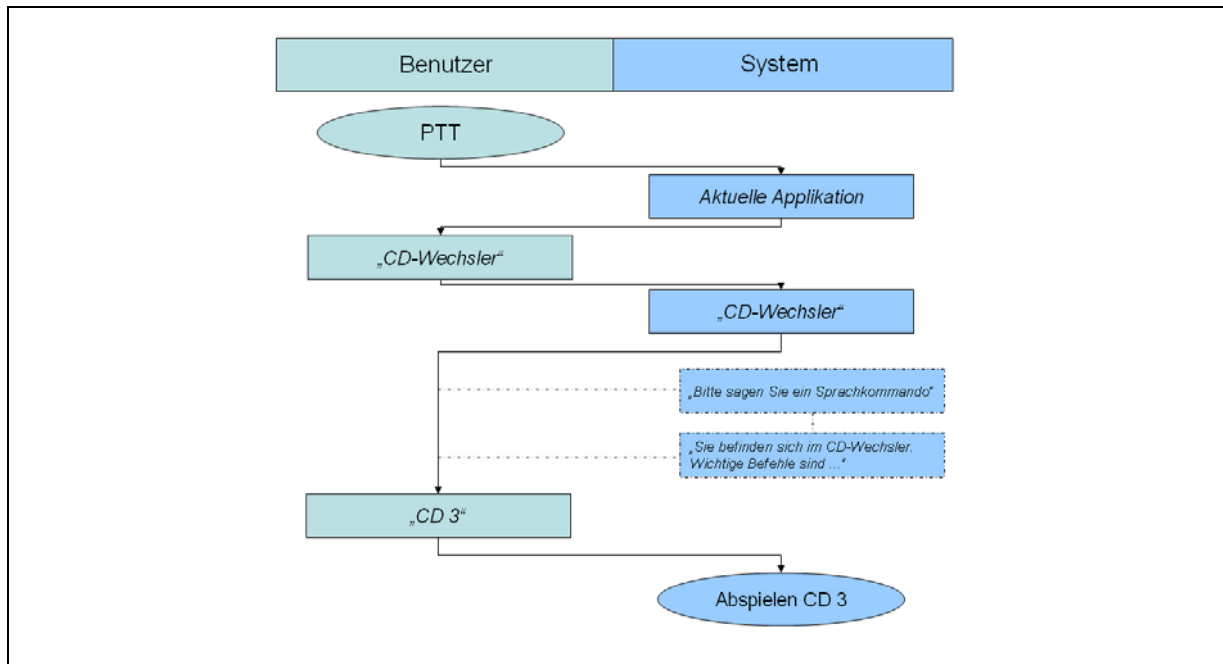


Abbildung 4.3: Bediensequenz „Abspielen CD 3“

Der Druck des PTT-Knopfs dient hierbei dem *Öffnen* des Mikrofons. Das System meldet daraufhin, auf welche Applikation der Benutzer in diesem Moment zugreifen kann. Mit den globalen Sprachkommandos, also jenen, die zu jedem Zeitpunkt des Dialogs zur Verfügung stehen, kann der Nutzer nun einen Applikationswechsel vollziehen: „Radio“, „Telefon“, „Navigation“ oder in diesem Falle „CD-Wechsler“. Nach erfolgreicher Spracherkennung meldet das System wieder die aktuell aktive Applikation. Anschließend signalisiert das System, dass es auf eine weitere Eingabe wartet. Der Anwender kann dann mit einem kontextabhängigen Befehl auf die Applikation zugreifen. Im CD-Wechsler löst die Eingabe des Sprachkommandos „CD 3“ also das Starten des ersten Liedes der CD im dritten Schacht aus. Wird dieser Befehl erfolgreich erkannt, führt das System ihn aus und die Bediensequenz ist beendet. Die Nutzereingabe „CD 3“ ist also ein Terminalbefehl. Andere Applikationen, insbesondere das Navigationssystem, sind in mehreren Schritten zu bedienen. Aktiviert der Benutzer das Navigationssystem, kann er z.B. „Stadt eingeben“ wählen. Daraufhin fordert das System ihn auf: „Bitte Stadt eingeben!“ Und öffnet automatisch das Mikrofon zur weiteren Eingabe. Nachdem dann ein Stadtname (z.B. „Hildesheim“) gesprochen wurde, bittet die Applikation um eine Bestätigung: „Sagten Sie Hildesheim?“ und wartet erneut auf eine Eingabe. Solch eine Sequenz setzt sich

fort bis ein Terminalbefehl erreicht ist. Der Nutzer muss also nicht vor jeder Eingabe wieder den PTT-Knopf drücken, sondern wird durch einen Dialog geführt. Ist eine Eingabe einmal nicht korrekt oder wird falsch verstanden, fordert das System den Benutzer auf: „*Bitte geben Sie ein Sprachkommando ein!*“ Nach der dritten fehlgeschlagenen Eingabe, wird der Dialog abgebrochen und zu dem aktuell aktiven Gerät zurückgegangen.

Der *SDS Demonstrator* bildet also die Dialogstruktur und die Bedienung der einzelnen Geräte eines Fahrerinformationssystems komplett ab. Dabei liegt allerdings der Fokus deutlich auf der Abbildung der Struktur des Sprachdialogs. Gerade die grafische Darstellung der Geräte ist demnach in der Evaluierung zu vernachlässigen. Wohingegen die Auflistung der Sprachkommandos, die Führung durch den Dialog, Spracherkennung- und -ausgabe den tatsächlichen Entwicklungsstand der Sprachbedienung widerspiegeln.

#### **4.1.2. „Helen“ – Die persönliche Informationsagentin**

Das zweite System dieser vergleichenden Evaluierung unterscheidet sich nicht von der Funktionsweise des *SDS Demonstrator*. Die zugrunde liegende VisualBasic-Anwendung entspricht der des ersten Systems. Lediglich die Fenster- und Buttonanordnung bzw. -benennung ist leicht unterschiedlich. Die Funktionalitäten und die Benutzerführung bleiben aber gleich. Der grundlegende Unterschied ist *Helen – Die persönliche Informationsagentin*. Das Kernelement des zweiten Prototyps (s. *Abb. 4.4*) ist ein virtueller Charakter der Charamel GmbH.

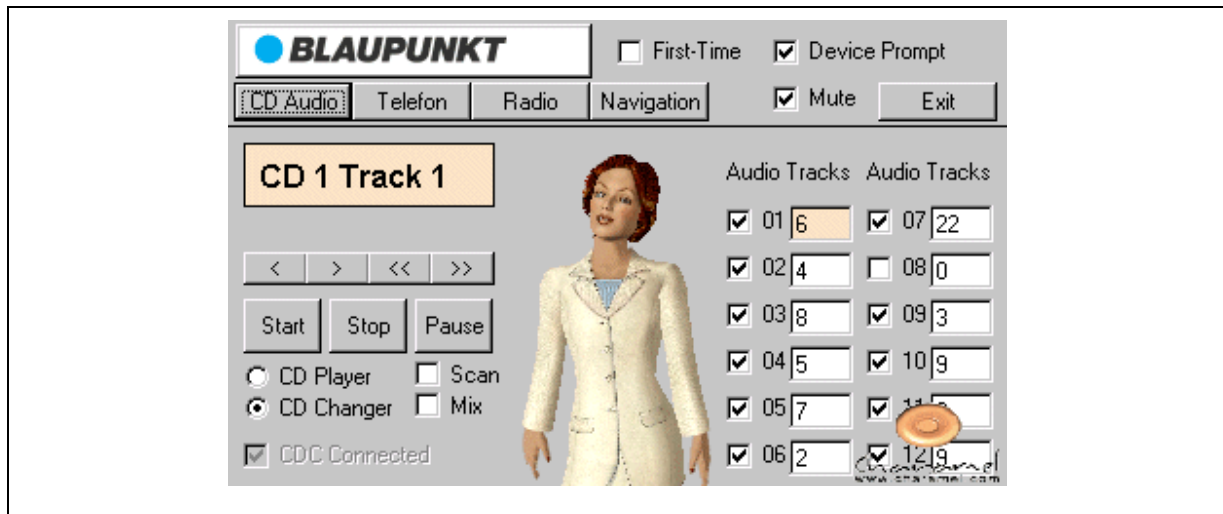


Abbildung 4.4: Prototyp *Helen*

Dieses Unternehmen wurde von der Blaupunkt GmbH beauftragt, das vorliegende Sprachdialogsystem mit einem menschlichen Antlitz zu versehen. Bereits in Systemen anderer Unternehmen wurden die virtuellen Charaktere der Charamel GmbH als verkaufsfördernde Maßnahme eingesetzt. So bestand eine Auswahl an Charakteren und Darstellungsmöglichkeiten für den Einsatz im *SDS Demonstrator*. Der Charakter *Helen* wurde ausgewählt, in einem Overlay<sup>7</sup> über die bestehende Anwendung gelegt und an die Sprachausgabe angebunden. Die 3D-Animation basiert auf der CharActor™-Software der Charamel GmbH. Dieses Tool übernimmt das Rendering<sup>8</sup> und die Steuerung der in Echtzeit gerechneten Bewegungen. Dabei greifen definierte Steuerzeichen innerhalb des Dialogtextes, der an die TTS-Software gesendet wird, auf ein bestehendes Bewegungsregister zurück, das somit parallel zum ausgegebenen Sprachsignal abgerufen werden kann. Die TTS-Technologie der Firma Nuance (ehemals Scansoft), die bei beiden Prototypen für die Generierung der Sprachausgabe auf Basis von Dialogtexten verantwortlich ist, kann zusätzliche Informationen zu Visemen<sup>9</sup> und Länge der Ausgabe an das CharActor™-Modul übertragen und somit für Lippensynchronität bzw. zeitgenaue Bewegungsabläufe

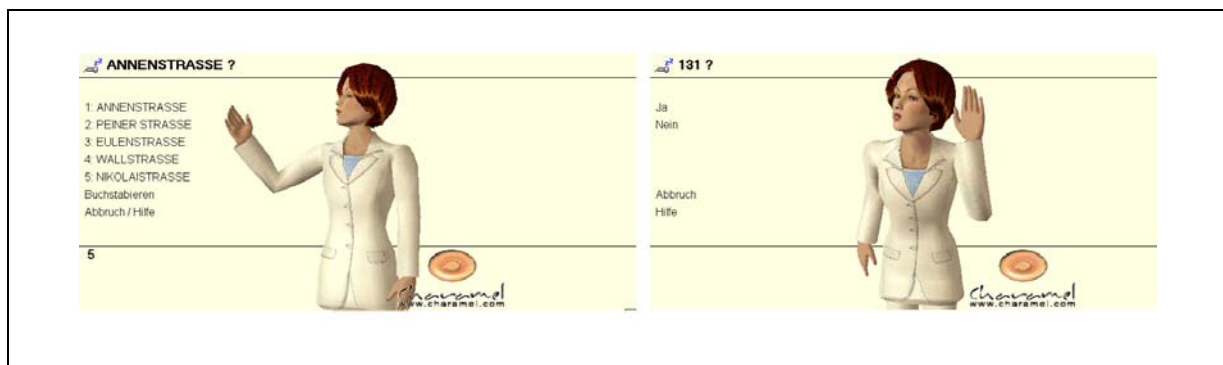
<sup>7</sup> Die Bezeichnung Overlay meint eine grafische Schicht mit durchlässigem Hintergrund, die es ermöglicht Inhalte über bestehende Anwendungsfenster einzublenden.

<sup>8</sup> „Im Bereich der Computergrafik bezeichnet Rendern die Erzeugung eines digitalen Bildes aus einer Bildbeschreibung. Bei einer zweidimensionalen Beschreibung ist damit die Rasterung gemeint. Bei 3D-Szenen sind räumliche Objekt-Daten Teil der Bild- oder Szenenbeschreibung. Die Software dazu wird als (3D-)Renderer bezeichnet.“ (URL: WIKIMEDIA1)

<sup>9</sup> visuelle Phoneme; also Informationen zur visuellen Ausprägung der Lippenbewegung



sorgen. Gegenüber dem Anwender tritt diese Figur also als menschliche Repräsentation des Sprachdialogs auf. *Helen* verhilft der synthetischen Stimme des Fahrerinformationssystems zu einer menschlichen Verkörperung und ergänzt die gesprochene Sprache mit menschenähnlichen nonverbalen Informationen. So verwendet sie z.B. entsprechende Deiktik, um auf die Liste der zur Verfügung stehenden Sprachkommandos zu verweisen. (Abb. 4.5 links)



**Abbildung 4.5: Beispiele der Körpersprache des AIA *Helen***

Weitere Elemente ihrer Körpersprache sind Gesten, die die Sprachhandlung zusätzlich unterstützen wie bei der Nachfrage: „Sagten Sie 131?“ (Abb. 4.5 rechts) *Helen* agiert also somit als eine Moderatorin, die durch die Bediensequenzen des Fahrerinformationssystems führt. Ist die Programmierung des Navigationsgeräts abgeschlossen, so verabschiedet sie sich vom Bildschirm und gibt den Blick auf die Kartennavigation frei. Bisher ist das Ausblenden *Helens* jedoch nur bei Beginn der Zielführung implementiert. Nach anderen Terminalbefehlen bleibt sie weiterhin sichtbar.

Die Wahl der Erscheinung hing dabei unmittelbar von der in Blaupunkt-Produkten verwendeten TTS-Stimme ab. Aus dem Charakter-Inventar der Charamel GmbH wurde eine Dame in Geschäftskleidung gewählt. Dies entspricht einem Charakter, der mit der klaren, weiblichen Stimme der Blaupunkt Navigationssysteme und der Aufgabe, einer Informationsagentin, im Einklang steht.

Diese beiden Prototypen sprachbedienbarer Fahrerinformationssysteme stellen die Untersuchungsgegenstände dieser Evaluierung dar. In welcher Weise diese Systeme untersucht wurden, beschreibt der folgende Abschnitt.

## **4.2. Methodenwahl**

Die in *Kapitel 2.2* vorgestellten Methoden und ihre Unterscheidung in Experten- und Benutzerorientierung haben verdeutlicht, dass sowohl eine analytische als auch eine empirische Betrachtung nötig sind, um eine umfassende Bewertung der Usability eines Systems zu erlangen. Für die vorliegende Untersuchung von sprachbedienbaren Fahrerinformationssystemen waren für die Methodenwahl folgende Parameter entscheidend:

- Es lagen zwei funktionsfähige Prototypen vor.
- Die Systeme sollten im Vergleich zueinander analysiert werden.
- Die Bedienung des Systems hatte per Sprache zu erfolgen.
- Auf Seiten der Blaupunkt GmbH wie auch der Universität Hildesheim standen Usability-Experten für eine analytische Untersuchung zur Verfügung.
- Die Untersuchung befasst sich einerseits mit der Usability bezogen auf Effektivität und Effizienz, aber andererseits auch mit dem Aspekt der Benutzerakzeptanz und Emotionalität.

Diese Kriterien führen zu einer 2-stufigen Vorgehensweise. Im ersten Schritt bietet sich das vorhandene Expertenwissen der Blaupunkt GmbH (bzw. der Bosch GmbH) und der Universität Hildesheim an, um das Sprachdialogsystem bezüglich anerkannter Gestaltungsrichtlinien zu analysieren. Es wurde also zunächst der Prototyp *SDS* dahingehend untersucht, inwiefern dieser bereits die erweiterten Usability-Richtlinien für sprachbedienbare Fahrerinformationssysteme erfüllt. Die Methodik der heuristischen Evaluierung erlaubte anhand von vorgegebenen Kriterien eine Bestandsaufnahme der Bedienbarkeit des Systems aus Expertensicht und Benutzersicht (*Cognitive Walkthrough*) in Expertengesprächen vorzunehmen, um dann in gemeinsamen Gruppendiskussionen das Referenzsystem *Helen* gerade bezüglich der gefundenen Mängel auf potenzielle Vor- bzw. Nachteile zu untersuchen. Mit der daraus resultierenden Erkenntnis über eventuelle Usability-Mängel und potenzielle Auswirkungen des anthropomorphen Prototyps konnte gezielt ein Benutzertest konzipiert werden. Der Benutzertest bediente sich der klassischen Methoden eines Usability-Tests. Die oben genannten Parameter schlossen jedoch

einige Methoden wie beispielsweise das *laute Denken* aus. Spricht der Benutzer während der Bedienung über seine Beobachtungen und Intention, so verhindert dies eine fehlerfreie Spracherkennung des Systems. Es wurde also ein kontrollierter Benutzertest konzipiert, der es ermöglichte einerseits messbare Daten über die effiziente und effektive Bedienbarkeit der Systeme zu erheben und andererseits die emotionale Reaktion der Nutzer auf die Benutzerschnittstellen und deren Akzeptanz zu messen. Die genaue Umsetzung dieses Benutzertests wird im *Abschnitt 4.4.* beschrieben.

### **4.3. Heuristische Evaluierung der Prototypen**

Die erste Stufe dieser Untersuchung ist also die heuristische Evaluierung. Sie dient einer ersten Einschätzung der vorliegenden Prototypen. Hierfür erstellte Heuristiken ermöglichen eine professionelle Bewertung der allgemeinen Usability der beiden Systeme und ihrer Schwachstellen. Die geführten Experteninterviews sollen zudem eine Einschätzung geben, welche Auswirkungen der Einsatz eines anthropomorphen Interface-Agenten auf die Bedienung der Systeme hat. Ausgehend von dieser ersten professionellen Erkenntnis kann man konkrete Erwartungen an den Benutzertest formulieren und zudem erkannte Mängel bzw. kritische Stellen in der Aufgabenstellung der Benutzerstudie gezielt ansteuern oder vermeiden.

#### **4.3.1. Konzeption und Durchführung der Experten-Evaluation**

Für diese Evaluierung wurden fünf Experten ausgewählt. Auf der einen Seite waren dies zwei wissenschaftliche Mitarbeiter des Instituts für angewandte Informationswissenschaften der Universität Hildesheim. Auf der anderen Seite wurden drei Mitarbeiter der Bosch GmbH aus der Abteilung *CR/AEH3 – Corporate Research Human Machine Interaction* für die heuristische Untersuchung herangezogen. Die Expertise der Teilnehmer lag somit primär in der Erforschung der Mensch-Maschine-Interaktion. Die notwendige Distanz zu Blaupunkt-Produkten oder sogar den zu untersuchenden Systemen war durch diese Auswahl gewährleistet. Die

Heuristiken, anhand derer die Systeme analysiert wurden, basieren auf einem allgemeinen Ansatz von Rolf Molich und Jakob Nielsen (1990) bzw. Werner Schweibenz und Frank Thissen (2002). Diese Heuristiken stellen eine Art allgemeines Leitbild für Benutzerschnittstellen dar, die in Wissenschaft und Praxis weit verbreitet sind. Auch den in dieser Studie beteiligten Experten waren diese Prinzipien bekannt, und so bot sich eine Orientierung an dieser Arbeit an, jedoch war eine Anpassung an sprachdialogspezifische Merkmale notwendig. Zusätzlich dazu wurde eine weitere Heuristik aufgenommen, die sich dem Sachverhalt des bereits erwähnten Aspekts der Emotionalität und Motivation zuwendet. Die verwendeten Heuristiken formulierten demnach Design-Prinzipien für die folgenden Kategorien<sup>10</sup>:

- 1 - Sichtbarkeit des Systemstatus
  - 2 – Übereinstimmung zwischen dem System und der realen Welt
  - 3 – Benutzerkontrolle und –freiheit
  - 4 – Konsistenz und Standards
  - 5 – Transparenz
  - 6 – Flexibilität und Effizienz der Benutzung
  - 7 – Hilfe beim Erkennen, Diagnostizieren und Beheben von Fehlern
  - 8 – Hilfe und Dokumentation
  - 9 – Motivation und Unterstützung
- (in Anlehnung an: SCHWEIBENZ / THISSEN 2002: 101ff.)

Die Gespräche mit den jeweiligen Teilnehmern wurden einzeln geführt und dauerten 90 Minuten. Die Experten wurden zunächst gebeten, die Heuristiken durchzulesen und sich mit den dort erwähnten Kategorien vertraut zu machen. Anschließend wurden Fragen zu Formulierungen oder zur Vorgehensweise geklärt. Dann bekam der Experte die Gelegenheit, das erste System<sup>11</sup> in freiem Browsing kennen zu lernen. Angewiesen durch den Versuchsleiter wurden daraufhin 45 Minuten lang verschiedene Funktionen untersucht und Aufgabenschritte vollzogen. Zu jeder Zeit war der Experte dazu angehalten, empfundene Abweichungen zu den in den Heuristiken formulierten Idealen zu äußern. Der Versuchsleiter notierte die genannten Mängel und bat zusätzlich den Teilnehmer, sich ausführliche Notizen zu machen und die genannten Punkte einer Kategorie zuzuordnen.

---

<sup>10</sup> Für die vollständige Definition der Heuristiken s. *Anhang A*

<sup>11</sup> Die Reihenfolge der Systeme wurde von Teilnehmer zu Teilnehmer variiert.

Nach dieser Sammlungsphase von Usability Mängeln, füllte der Experte einen Fragebogen zur abschließenden Gesamtbewertung aus. Dieser Fragebogen (s. CD-ROM) basiert ebenfalls, wie auch die Heuristiken, auf internationalen Standards der Software-Ergonomie. Von Prümper und Anft 1993 entwickelt, eignet sich „der Fragebogen ISONORM 9241/10 [...] zur Beurteilung bereits eingesetzter Software ebenso wie zur Beurteilung von Prototypen. Der Fragebogen liefert erste Hinweise auf ergonomische Schwachstellen von Softwaresystemen.“ (URL: ERGO-ONLINE) Die Beantwortung der an diesen Fragebogen angelehnten Kategorien erforderte nur geringen Zeitaufwand und sollte den Experten neben einer analytischen Meinung auch zu einer System-Gesamtbewertung bewegen, die seine Eindrücke auf einer Fünfer-Skala festlegt. Diese betraf die nach der ISO Norm 9241/10 definierten Items: *Aufgabenangemessenheit*, *Selbstbeschreibungsfähigkeit*, *Steuerbarkeit*, *Erwartungskonformität*, *Fehlertoleranz* und *Lernförderlichkeit*. Vor dem Hintergrund der potenziellen Auswirkungen des Einsatzes eines anthropomorphen Interface-Agenten wurde zusätzlich ein weiteres Item ergänzt: *Motivation*. In Anlehnung an Shneiderman erfragte dieses Item ihre Einschätzung zur Unterstützung und Motivation bzw. die allgemeinen Empfindungen des Teilnehmers während der Benutzung des Systems. (vgl. 1998: 136) Dies sollte Erkenntnisse darüber liefern, ob die Gestaltung der Benutzungsschnittstelle einen Einfluss auf die Gemütslage des Benutzers hat. Die schriftliche Bewertung der Systeme in Form der Fragebögen sollte somit einen ersten Fingerzeig geben, welche subjektiven Unterschiede zwischen den beiden Systemen bezüglich der Usability und empfundenen Emotionalität von den Teilnehmern des Benutzertests zu erwarten sein könnten.

Darauf folgte dann eine 30-minütige Inspektion des zweiten Systems. In dieser Zeit sollten eventuelle Verbesserungspotenziale gegenüber den vorher genannten Fehlern des ersten Systems analysiert werden. Um eine eventuelle Verfälschung der Eindrücke zu vermeiden, wurde die Reihenfolge der Systeme von Experte zu Experte abgewechselt. Ergebnis dieser Einzelsitzungen war schließlich eine umfangreiche Liste mit Abweichungen der untersuchten Systeme von den Heuristiken. Um die einzelnen Punkte in ihrer Schwere zu bewerten, fanden gemeinsame Expertendiskussionen statt. Eine mit den beiden Experten der Universität und eine

mit den Bosch-Mitarbeitern. Sie dienten der Zusammenführung der Erkenntnisse und der gemeinsamen Bewertung der erkannten Mängel. Das sog. *Severity Rating* bot den Teilnehmern die Gelegenheit, nach eigener Einschätzung die gefundenen Fehler zu kategorisieren. Ihnen wurde dazu eine Skala (s. *Tabelle 4.1*) zur Verfügung gestellt, welche die jeweilige Bedeutung eines Fehlers bewerten sollte. Die Einordnung eines Aspekts in die entsprechende Kategorie erfolgte in der gemeinsamen Absprache aller Teilnehmer.

0 - Kein Usability Problem
1 - Kosmetisches Problem
2 - Geringfügiges Problem
3 – Bedeutendes Problem
4 – Usability Katastrophe

**Tabelle 4.1: Severity Rating der Experten-Evaluation**

(vgl. URL: NIELSEN)

Zusätzlich wurde mit den Teilnehmern gemeinsam das Potenzial des Einsatzes eines anthropomorphen Interface-Agenten, wie es im System *Helen* der Fall ist, frei diskutiert. Dabei wurde zum einen die mögliche Verbesserung der Usability als aber auch die zu erwartende Nutzerakzeptanz analysiert.

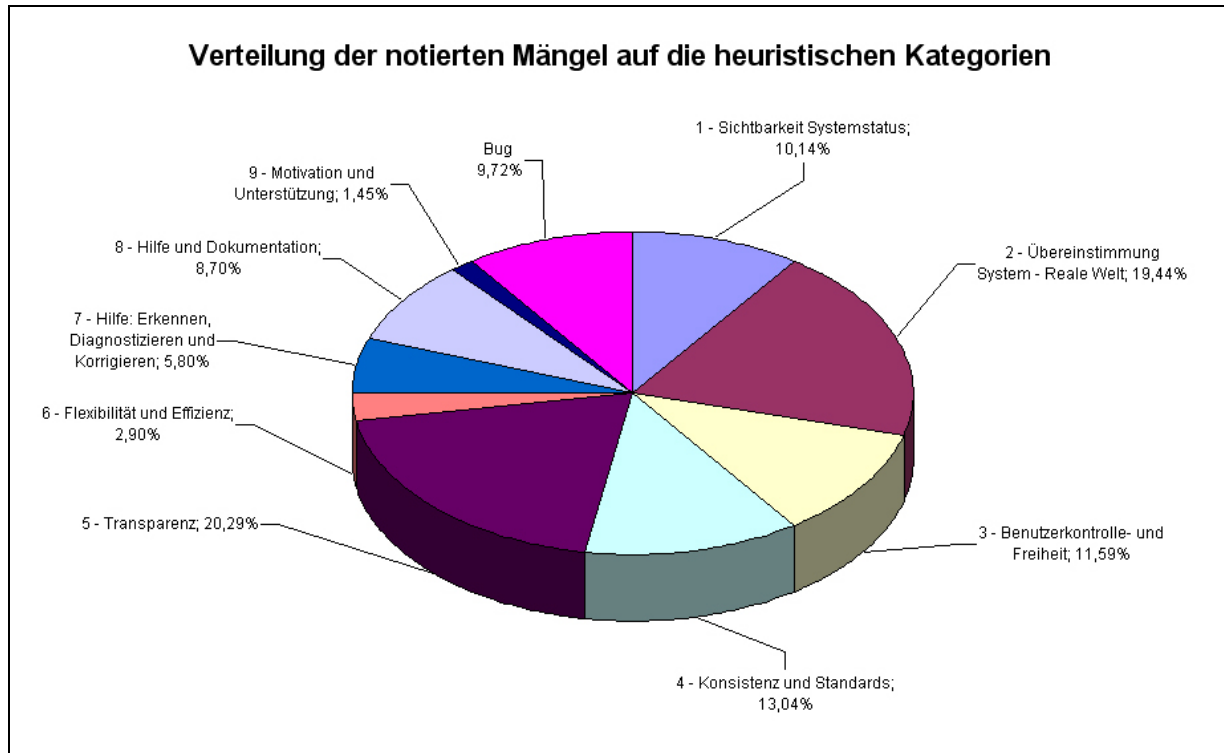
### 4.3.2. Ergebnisse

Die fünf Teilnehmer notierten insgesamt 92 Punkte, nach denen das von ihnen untersuchte System von den Usability-Heuristiken abwich. Dabei ist zu bemerken, dass in einer Art *Brainstorming* der Versuchsleiter keinerlei Vorsortierung der bemerkten Aspekte vornahm. Zunächst galt es alle Bemerkungen der Teilnehmer zu notieren. Bis auf einen Experten, der 10 Mängel notierte, lag hierbei das Mittel der gefundenen Usability-Mängel pro Experte bei 21. Die so entstandene Liste mit insgesamt 92 *Token*<sup>12</sup> beinhaltete 23 identische oder sehr ähnliche Punkte.

---

<sup>12</sup> „Die Type-Token-Relation ist eine aus der Statistik übernommene Terminologie zur Unterscheidung zwischen einzelnen sprachlichen Äußerungen ( = token) und der Klasse der diesen Äußerungen zugrunde liegenden abstrakten Einheiten ( = types).“ (URL: UNIVERSITÄT TRIER)

Insgesamt ergaben die Expertengespräche also eine Liste mit 69 *Types*. Die von den Experten selbst vorgenommene Einteilung in die heuristischen Fehlerkategorien ergab folgende Verteilung (s. Abb. 4.6):



**Abbildung 4.6: Verteilung der notierten Mängel auf die heuristischen Kategorien**

Hierbei lässt sich feststellen, dass gerade die Kategorien 2 – *Übereinstimmung System – Reale Welt* und 5 – *Transparenz* die meisten Anteile der gefundenen Usability-Mängel beinhalten. Diese Kategorien enthalten vielfach Fehler betreffend der Terminologie, wie z.B.:

#### Kategorie 2 – Übereinstimmung System – Reale Welt

*ID 222: Terminologie: "CD-Wechsler" "CD-Audio" "CD-Player" "CD-Changer" unverständlich.*

*ID 106: CD: Beschriftung "Lied <1..30>/CD <1..12>" irreführend: falsche Reihenfolge und kryptische Darstellung verwirren Nutzer bei kurzem Blick auf Auswahl.*

oder

#### Kategorie 5– Transparenz

*ID 314: Bei „Buchstabieren“ fordert das System auf: „Bitte geben Sie ein Sprachkommando ein.“*

ID 122:                      NAVI: „Routenoptionen ändern" wird in der Hilfe vorgelesen erscheint aber nicht auf der grafischen Oberfläche.

Diese Arten von Fehlern sind stark abhängig von der Gestaltung des Prototyps. Sie sind somit weniger schwer zu gewichten, denn man geht davon aus, dass Button-Bezeichnungen und einige Teile der Sprachausgabe bei dieser Art von Prototyp nicht bis ins letzte Detail auf Konsistenz bzw. Transparenz geprüft wurden. Einige Missstände in der Systemnavigation, wie plötzliche Abbrüche, wurden zudem als reine Programmfehler (*Bug*) eingestuft, da das Systemverhalten kein beabsichtigtes Konzept erkennen ließ. Solche Fehler sind also für eine allgemeine Usability-Bewertung zu vernachlässigen. Auch wenn Sie im anschließenden Severity-Rating also durchaus hoch eingestuft wurden, sind diese Mängel generell weniger besorgniserregend. Dennoch fanden viele solcher Anmerkungen ihren Niederschlag in den Ergebnislisten und dürfen bei einer späteren Produkt-Spezifikation nicht in Vergessenheit geraten. In der gemeinsamen Bewertung der notierten Punkte, gab es eine recht hohe Anzahl an *Usability Katastrophen* und *Bedeutenden Problemen* (s. Abb. 4.7).

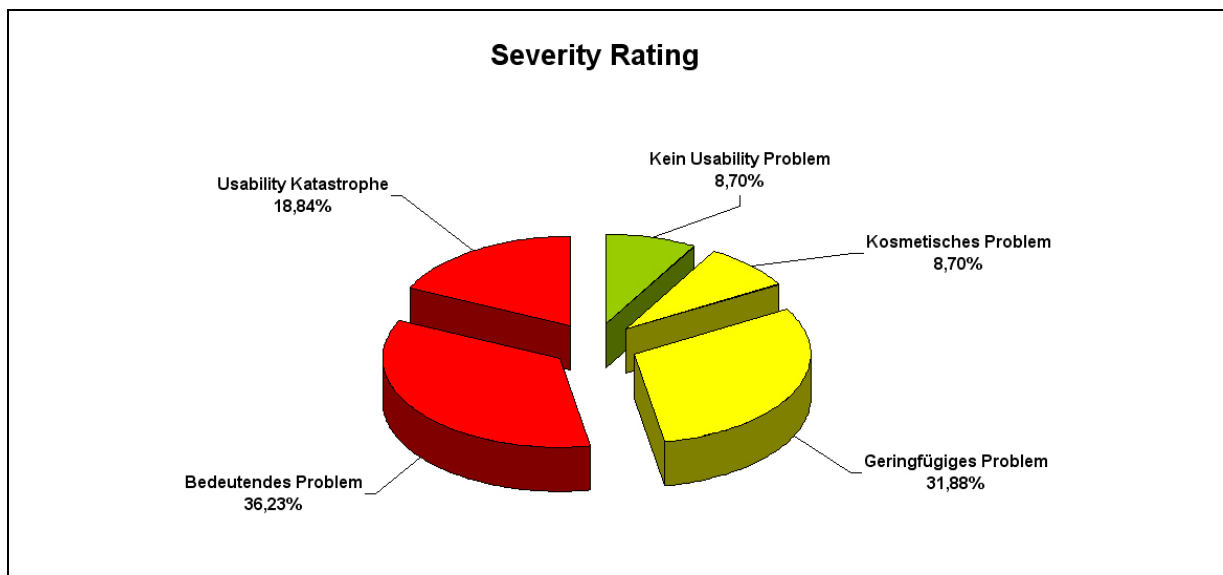


Abbildung 4.7: Severity-Rating der Experten-Evaluation

Dazu gehörten neben den oben genannten Prototyp-spezifischen Punkten aber auch software-ergonomische Bedenken, die in *Kapitel 2* genannte Prinzipien zur



benutzergerechten Gestaltung von Sprachdialogsystemen vernachlässigen. So sind beispielsweise folgende Mängel als *Usability-Katastrophen* bewertet worden:

Kategorie 3– Benutzerkontrolle- und Freiheit

- ID307: *Automatischer Abbruch nach 3 Eingabeversuchen bzw. nach 3 ausgelassenen Eingabemöglichkeiten bringt Nutzer ungewollt wieder zurück an Beginn der Eingabesequenz!*
- ID 217: *Fehlende Möglichkeit, in der Bediensequenz einen Schritt ZURÜCK zu gehen.*

Kategorie 5– Transparenz

- ID 309: *Befehlsliste ist erst sichtbar, wenn PTT gedrückt und System bereit für Eingabe. Eine vorherige Orientierung der zur Verfügung stehenden Befehle ist nicht möglich.*

Als *Bedeutende Mängel* wurden folgende Punkte bezeichnet:

Kategorie 1– Sichtbarkeit Systemstatus

- ID 304: *Der Befehl "Radio" führt direkt zur Unterauswahl ohne dass das Gerät vorher angezeigt wurde und somit bekannt ist welcher Sender aktiv ist oder welche Sender zur Auswahl stehen.*
- ID 202: *Die lange Verzögerung zwischen Knopfdruck PTT und Systembereitschaft stellt eine große Gefahr für den Dialogfluss und die fehlerfreie Eingabe des Users dar.*

Kategorie 2– Sichtbarkeit Systemstatus – reale Welt

- ID 121: *NAVI: Der Befehl "Suche um Stadt" / "Suche in der Nähe" und ihre Funktionen sind nicht ersichtlich.*
- ID 010: *Radiosender kann man nicht mit Namen aufrufen. ("Jump", "NDR2")*

Kategorie 3– Benutzerkontrolle- und Freiheit

- ID 108: *TEL: " <0...9> und weiter?" ist eine zu direkte Frage, die den Benutzer dazu auffordert, eine Eingabe fortzusetzen, obwohl nicht offensichtlich ist, dass die Eingabe nicht schon beendet sein könnte. Eine offenerere Frage sollte gestellt werden.*

Kategorie 4– Konsistenz und Standards

- ID 107: *CD: "Anspielen" und "Mix an/aus" stehen zusammen sind aber inhaltlich keine zusammengehörigen Funktionen.*
- ID 110: *TEL: Das System gibt hier Feedback "Die Nr. wird gewählt", im CD-Wechsler fehlt allerdings solche eine Bestätigung wie z.B. "CD6 ausgewählt." Dies stört die Konsistenz.*

Kategorie 5– Transparenz

- ID 402: *Bei Telefonnummerneingabe ist nicht erkenntlich welches Format und wie viele Nummern eingegeben werden können.*
- ID 401: *"Globale Sprachkommandos" stehen nicht überall zur Verfügung. In Unterebenen der Applikationen ist ein Wechsel zu anderen Applikationen nicht möglich.*

Kategorie 6– Flexibilität und Effizienz

- ID 320: *Starre Kommandostruktur verhindert schnellen Durchlauf durch Eingabesequenz.*
- ID 413: *Buchstabierende Eingabe nicht von Beginn an möglich.*

Kategorie 7– Hilfe: Erkennen, Diagnostizieren und Korrigieren

- ID 221: *Wiederholt man bei Stadtbestätigung die Eingabe noch einmal, so verwendet das System die dann verstandene Stadt und leitet ohne Bestätigungsabfrage zu "falscher Stadt"*

Kategorie 7– Hilfe: Erkennen, Diagnostizieren und Korrigieren

- ID 007: *Die Hilfefunktion ist nicht sehr hilfreich, da nur die Menüoptionen vorgelesen werden. Die Erwartung des Benutzers ist eine detaillierte Erklärung der einzelnen Funktionen und ihrer Bedeutung bzw. eine Anleitung im Sinne von "Was möchten Sie tun?" oder eine Erläuterung der möglichen Eingabeformate etc.*

Die vollständige Auflistung der gefundenen Mängel und die Bewertung der Bedeutung für die Usability des Systems finden sich in *Anhang B*. Die oben genannten Beispiele sollen einen Eindruck von der analytischen Arbeit der Experten geben. Sie bewerteten einerseits die reine Erfüllung der Heuristiken, versuchten jedoch andererseits ebenso in einem *Cognitive Walkthrough* eventuelle Bedienschwierigkeiten der Benutzer bei der Bearbeitung bestimmter Aufgaben zu simulieren.

Im Anschluss an die heuristische Evaluierung gab jeder Teilnehmer seinen persönlichen Eindruck im Fragebogen wieder. Die Bewertung der einzelnen Usability-Aspekte der beiden Systeme *SDS* und *Helen* ergab dabei folgendes (s. *Abb. 4.8*):

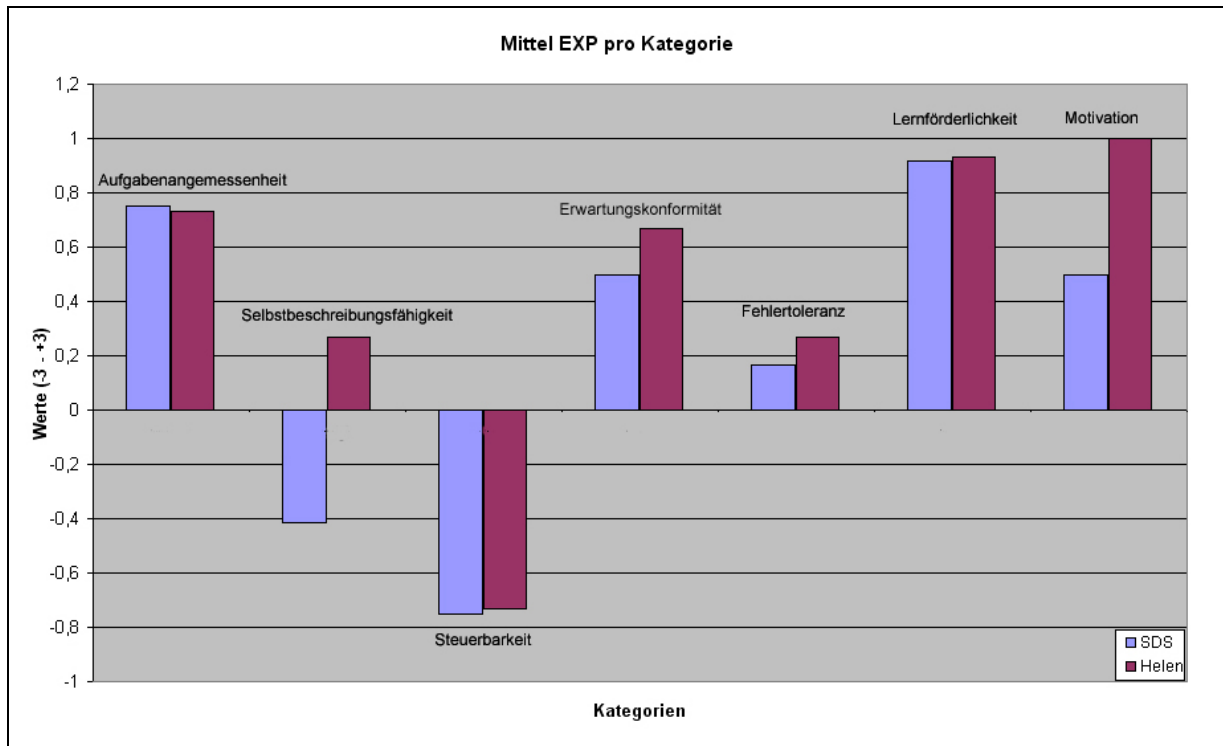


Abbildung 4.8: Fragebogenergebnisse der Experten: Usability im Vergleich

Eine signifikante Abweichung der Systembewertungen ist nicht zu beobachten.<sup>13</sup> Bei einer Befragung von 5 Teilnehmern war dies nicht zu erwarten. Die jeweiligen Varianzen der Stichprobe sind entsprechend hoch. Doch bei der Beantwortung mehrerer Unterpunkte der jeweiligen Usability-Items sind zwei entscheidende Abweichungen zu bemerken. Das Fahrerinformationssystem mit virtuellem Charakter schnitt in der Stichprobenbefragung deutlich besser in den Punkten *Selbstbeschreibungsfähigkeit* und *Motivation* ab. Auch in den Experten-Diskussionen wurde deutlich, dass die Teilnehmer davon ausgehen, dass der Einsatz des AIA *Helen* zur Attraktivität des Produkts beiträgt und somit zu einer höheren Nutzerzufriedenheit und größerem Verständnis der Funktionen und ihrer Inhalte beiträgt. Allerdings wurde gerade von den Experten aus der Automobilbranche sehr kritisch darauf hingewiesen, dass eine grafisch aufwändige Animation im Blickfeld des Fahrers ablenkend und somit sicherheitsgefährdend wirke. Von dem vermuteten

<sup>13</sup> Zur Berechnung der statistischen Relevanz wurde der Zweistichproben t-Test: unterschiedlicher Varianzen (Excel) angewandt. „Er bestimmt die Sicherheit eines Vergleichs von Mittelwerten zwischen zwei normalverteilten Zufallsgrößen.“ (PREIM 1999: 257) So kann man mit einer Irrtumswahrscheinlichkeit von  $p \leq 0,05$  feststellen, ob eine Abweichung von Messwerten beider Systeme statistisch signifikant ist und sich auf die Gesamtpopulation anwenden lässt.

Potenzial würde solch eine Anwendung also nur profitieren können, wenn die Implementierung sich an anerkannten Sicherheitsrichtlinien wie z.B. die *European Statement of Principles on the Design of Human Machine Interaction* der eSafety-Initiative der Europäischen Kommission (ESoP) orientiert.

Der kommende Abschnitt stellt die Auflistung der Ergebnisse der Experten-Evaluierung ins Licht der Untersuchungsperspektive und leitet mögliche Schlussfolgerungen und Fragestellungen für den Benutzertest ab.

### **4.3.3. Schlussfolgerungen der heuristischen Evaluierung**

Die umfangreiche Liste der Usability-Mängel des Sprachdialogsystems erfüllt zwei Zwecke. Sie ermöglicht einen Überblick über Verbesserungsmöglichkeiten der Dialogstruktur, des Oberflächendesigns und der technologischen Umsetzung. Diese Erkenntnisse sollen bei der Weiterentwicklung der Sprachtechnologie Level 3 der Blaupunkt GmbH Berücksichtigung finden und fließen in die Spezifikation möglicher Produktausprägungen ein. Die Erkenntnisse zur Usability des Sprachdialogsystems und Verbesserungsmöglichkeiten der reinen Dialoggestaltung sind zwar Teil der Entwicklungsarbeit der Blaupunkt GmbH, sollen aber nicht primär in dieser Arbeit behandelt werden. Sie dienen in dieser Studie vielmehr einem weiteren Zweck. Für die Konzeption eines kontrollierten Benutzertests ist es von enormer Bedeutung, sich der potenziellen Fehlerstellen innerhalb einer Dialogsequenz bewusst zu sein und diese entweder zu korrigieren, zu umgehen oder den Benutzer bewusst in komplexe Situationen des Fehlermanagements zu leiten. Zusätzlich zu der Mängelliste ergaben auch die subjektiven Bewertungen der Systeme erste Hinweise welche Nutzer-Reaktionen bei einer Befragung einer größeren Stichprobe zu erwarten sind. Daher lassen sich aus den Ergebnissen der ersten Evaluationsstufe folgende Schlussfolgerungen für den Versuchsaufbau des Benutzertests ableiten.

#### 4.3.3.1. Anpassung der Prototypen

Für die weitere Evaluierung der Prototypen waren zunächst die häufig genannten Mängel bezüglich der Konsistenz, Transparenz und Terminologie von Bedeutung. Diese Punkte stellen einen potenziellen Störfaktor für eine kontrollierte Beobachtung des Nutzerverhaltens dar. Die auf VBASIC basierenden Programme galt es also hinsichtlich dieser Fehler zu korrigieren. In einer kurzen Programmierungsphase wurden die Bezeichnungen von Buttons, die Beschriftungen der verschiedenen Gerätefunktionen und die Größe und Position der Programmfenster auf einen Standard gebracht. (s. Abb. 4.9) Dabei wurden einheitliche deutsche Bezeichnungen gewählt und die Oberfläche für eine Auflösung von 800 \* 600 in der Darstellung optimiert.

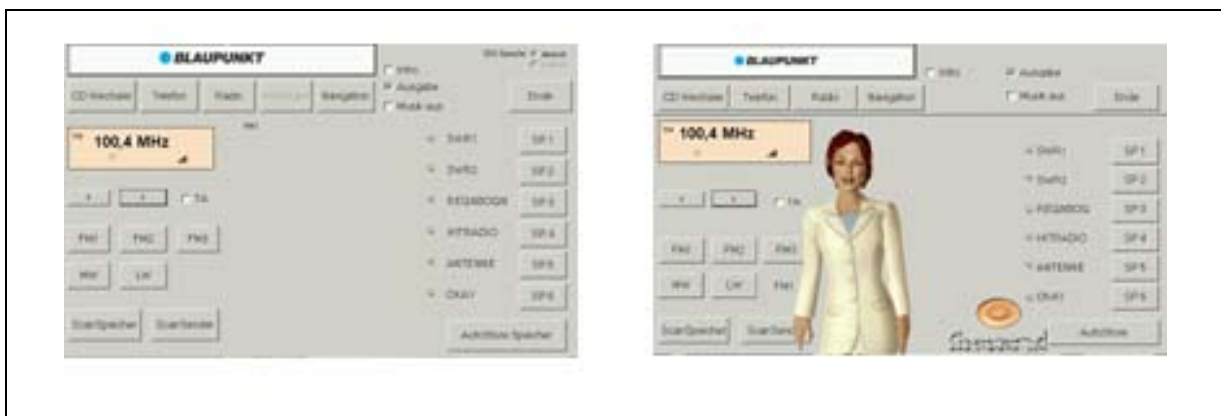


Abbildung 4.9: Angepasste Prototypen für den Benutzertest

Diese Art der Anpassung optimierte zwar nicht direkt die oben genannten grundlegenden Usability-Probleme der Dialogstruktur, sorgte jedoch für eine Minimierung des Störpotenzials in der Darstellung der Prototypen und erhöhte die Vergleichbarkeit der Oberflächen. Die erste Konsequenz aus der Experten-Evaluierung war also zunächst eine Anpassung der Prototypen. Die Systeme wurden so für den Benutzertest bezüglich des Versuchaufbaus optimiert.

#### 4.3.3.2. Hypothesen für den Benutzertest

Die große Anzahl der von den Experten analysierten Usability-Mängel weist darauf hin, dass bei der Beobachtung einer gesteuerten Aufgabenerledigung von Probanden mit beiden Prototypen ein gewisses Maß an Bedienfehlern und kritischen Situationen zu erwarten ist. Dies wird zusätzlich von der Tatsache gestützt, dass im *Severity-Rating* über 55% der Mängel eine hohe Bedeutung zugewiesen bekamen. Die negative Bewertung der Systemeigenschaft *Steuerbarkeit* im Experten-Fragebogen lässt erkennen, dass die Teilnehmer der heuristischen Evaluierung gerade in der Dialogstruktur und Benutzerführung noch Verbesserungsbedarf sehen. Zudem ist der hohe Anteil der Kategorien *Transparenz* und *Übereinstimmung System – reale Welt* an den erfassten Mängeln ein Indiz dafür, dass die zur Verfügung stehenden Funktionen und ihre Bedeutung nicht intuitiv erfassbar sind. Demnach müssten die Teilnehmer des Benutzertests Probleme haben, die richtigen Sprachkommandos für eine entsprechende Aufgabe zu erfassen. Zugegeben sind diese Erkenntnisse der Experten und Vermutungen bezüglich des Nutzerverhaltens teilweise sehr kritisch, dennoch hilft diese Kritik bei der Formulierung von Erwartungen an den Benutzertest. Bestätigt sich diese Bewertung auch im Usability-Test, dann ist mit einer hohen Fehlerrate und einem gewissen Frustrationsniveau der Probanden zu rechnen.

Gerade in den Kategorien *Selbstbeschreibungsfähigkeit* und *Motivation* haben die Experten den beiden Prototypen die größte Abweichung attestiert. In der subjektiven Bewertung der Systeme fiel das Urteil für das Dialogsystem mit der Agentin *Helen* deutlich positiver aus. Die Verkörperung des Sprachdialogs scheint also durch Verwendung der Körpersprache, Gestik und Mimik den Benutzer angemessener über die Systemfunktionalitäten zu informieren und zudem den Grad der Motivation zu steigern. Das Potenzial des Systems mit AIA sollte sich also auch in den Ergebnissen des Benutzertests niederschlagen.

Das Zusammenspiel der analysierten Usability-Mängel einerseits und der subjektiven Besserstellung des AIA andererseits in den angesprochenen Kategorien lässt für den Benutzertest folgende Arbeitshypothesen formulieren, die eine Erweiterung der allgemeinen Annahmen aus *Kapitel 3.3.* darstellen:

- *Arbeitshypothese AH 1*

Die Verwendung eines menschenähnlichen Charakters im Sprachdialog eines Fahrerinformationssystems bewirkt, dass im Vergleich zu einem System ohne AIA die Anzahl der Bedienfehler, und die Gesamtzeit zur Erledigung einer Aufgabe geringer bzw. der Grad der Aufgabenerledigung höher sind.

- *Arbeitshypothese AH 2*

Die qualitativ hochwertige Animation eines AIA bewirkt, dass die Teilnehmer des Benutzertests das System *Helen* bevorzugt in ihrem Fahrzeug einsetzen würden und eine höhere Ausgabebereitschaft signalisieren.

- *Arbeitshypothese AH 3*

Mängel bzgl. der Selbstbeschreibungsfähigkeit eines Systems führen zu Hilflosigkeit und Frustration des Benutzers. Der virtuelle Charakter jedoch vermag die empfundene Frustration zu senken bzw. die User-Motivation zu steigern.

- *Arbeitshypothese AH 4*

Das Potenzial des AIA kann nur ausgenutzt werden, wenn die Gestaltung der Interaktion mit dem virtuellen Charakter auf Sicherheitsrichtlinien im Fahrzeug abgestimmt ist und so eine empfundene Ablenkung von der Primäraufgabe des Fahrers vermeidet.

#### **4.3.3.3. Konzeption der Aufgabenstellung**

Als weitere Schlussfolgerung aus der ersten Evaluationsstufe galt es die Aufgaben, die die Probanden innerhalb des Usability-Tests erledigen sollten, so zu konzipieren, dass einerseits ein direkter Vergleich der messbaren und empfundenen Usability bzw. System-Akzeptanz möglich ist und andererseits die oben formulierten Hypothesen überprüft werden. Eine erste Einschätzung der Usability beider Prototypen aus Sicht der Experten ist also die Basis der Konzeption von Ablauf und Schwierigkeit der Aufgabenstellungen des Benutzertests. Geführte Aufgabenszenarien in einem Benutzertest sollen so, vergleichbare Daten zur Effektivität und Effizienz der Systeme erheben, sowie subjektive Bewertungen und messbare Empfindungen der Benutzer einholen.

Es wurden zwei unterschiedliche Aufgabenzettel (s. *Anhang C*) konzipiert, die jeweils durch ein Situationsszenario führten und dem Probanden Anweisungen zu den

verschiedenen Funktionalitäten der Prototypen geben. Da die Benutzer die Systeme ohne detaillierte Vorerfahrung nutzen sollten, wurde zunächst eine leichte Einstiegsphase gewählt. Mit leicht lösbaren Aufgaben konnte sich so der Proband an Lautstärke, Bedienkonzept und Anmutung des Systems gewöhnen. In der mittleren Aufgabenphase eines jeden Durchgangs wurde bewusst eine Funktionalität angesteuert, die den Benutzer in Schwierigkeiten bringen sollte. Hierfür wurden aus den Erkenntnissen der Experten-Evaluierung zwei Schlüsselstellen ausgesucht. Die langwierige Eingabesequenz zur Programmierung des Navigationssystems für eine komplette Anschrift stellte für Erstnutzer bereits eine Herausforderung dar. In diesem Fall wurden jedoch Straßenname und Hausnummer bewusst so gewählt, dass eine Fehlerkennung des Systems wahrscheinlich war. So ist die *Kirchstrasse* in ihrer Aussprache sehr sprecherabhängig. Ein Vortest mit zwei Testpersonen ergab, dass je nach Dialekt bzw. Deutlichkeit der Aussprache häufig *Königstr.*, *Föhrstr.* oder *Kimmichstr.* erkannt wurden. Auch die Hausnummer wurde so gewählt, dass der Nutzer in eventuelle Schwierigkeiten kommen konnte. Bei der Eingabe der Zahl 113 ist die deutliche Aussprache „*ehnhundertunddreizehn*“ für eine korrekte Spracherkennung notwendig. Auch hier war also ein gewisses Fehlerpotenzial zu erwarten. Im zweiten Szenario wurde diese kritische Situation bei einer anderen Funktionalität erzeugt. Alle Experten hatten einhellig darauf hingewiesen, dass der fehlende Hinweis auf mögliche Eingabeformate und die niedrige Erkennungsrate bei zu langer Zifferneingabe die Wahl einer Telefonnummer äußerst schwierig gestaltet. Zudem greift die Hilfefunktion gerade an dieser Stelle am wenigsten, denn der Hinweis „*Bitte geben Sie die Telefonnummer ein!*“ bietet dem Nutzer keine konstruktiven Lösungsvorschläge. Die Probanden sollten somit bei der Bedienung beider Systeme jeweils gezielt mit einer kritischen Situation konfrontiert werden, um gerade ihr Frustrationsniveau und das Krisenmanagement und mögliche systemspezifische Unterschiede zu analysieren. Die Beobachtung der Reaktionen in dieser Phase des Benutzertests lässt eventuelle Unterschiede der emotionalen Auseinandersetzung mit den Systemen und somit das Motivationspotenzial eines anthropomorphen Interface-Agenten zum Vorschein kommen. Am Ende der einzelnen Szenarien stand eine eher leicht lösbare Aufgabe. Einerseits die Programmierung des Navigationssystems auf eine beliebige Stadt und andererseits



das Einstellen einer geeigneten Kartenansicht stellten für den Benutzer abschließende Tätigkeiten dar, die eine eventuelle vorherige Frustration lösen und so beruhigen sollten.

Die Abfolge der beiden Aufgabenstellungen war bei jedem Probanden gleich, damit vergleichbare Daten gewonnen werden konnten. Allerdings war gerade im Erstkontakt mit sprachbedienbaren Dialogsystemen ein starker Lerneffekt zu erwarten. Um eine konstante Besserstellung des zweiten Systems abzufangen, wurde der Benutzertest also so konzipiert, dass eine Hälfte der Teilnehmer mit dem *SDS-Demonstrator* begann, während die andere Hälfte das System *Helen* verwendete, um die erste Aufgabenstellung zu bearbeiten. So war garantiert, dass beide Systeme mit ausgeglichenem Erfahrungsmaß der Benutzer bewertet wurden.

#### **4.4. Vergleichender Benutzertest der Prototypen**

Wie reagieren nun die potenziellen Nutzer des sprachbedienbaren Fahrerinformationssystems der Blaupunkt GmbH tatsächlich auf den virtuellen Charakter *Helen* und was für einen Einfluss auf Usability und Nutzerakzeptanz hat diese Figur? Diese zentrale Frage stellt neben den vorher genannten Hypothesen und Erwartungen den Ausgangspunkt dieser Studie dar. Die Antwort darauf gibt erst die Beobachtung tatsächlicher Erstkontakte von Anwendern mit den Prototypen. Die im vorherigen Kapitel hergeleiteten Hypothesen und konzipierten Aufgabenstellungen sollen helfen, dieser Untersuchung die nötigen Messparameter zu hinterlegen. Dabei orientieren sich die Testumgebung und die konzipierten Aufgabenstellungen an den von Preim geforderten Qualitätskriterien für Benutzerstudien:

- *Credibility (interne Validität)*: „ein glaubwürdiges Experiment ist ein sorgfältig geplantes und ausgewertetes Experiment.“<sup>14</sup>
- *Transferability (externe Validität)*: „ein wichtiger Aspekt der Übertragbarkeit besteht darin, dass die ausgewählten Aufgaben realistische und praktisch bedeutsame Aufgaben darstellen.“

---

<sup>14</sup> In Bezug auf die Glaubwürdigkeit eines Experiments erwähnt Preim auch die Bedeutung der ausgeglichenen Probandenauswahl hinsichtlich Alter, Geschlecht und Vorkenntnissen und die Aufteilung der Probanden in äquivalente Gruppen bei einer vergleichenden Evaluierung

- *Reliability*: „eine Studie ist zuverlässig, wenn die Anzahl und Auswahl der Testpersonen sowie die Dauer des Tests erwarten lässt, dass die Ergebnisse sich bei einer größeren Anzahl von Personen oder längeren Testdauer nicht wesentlich verändern würden.“ (1999: 254)

Im folgenden Abschnitt werden also unter Berücksichtigung dieser Kriterien der Versuchsaufbau und der Ablauf des Benutzertests beschrieben. Anschließend werden die verschiedenen Ergebnisse dargestellt, bevor dann in *Kapitel 5* diese Daten interpretiert und ausgewertet werden können.

#### **4.4.1. Probandenauswahl**

Im Vorfeld des Benutzertests galt es Probanden für die Untersuchung zu gewinnen. Dabei lag die Schwierigkeit zwar auch in der allgemeinen Suche nach Freiwilligen, aber viel mehr noch in der korrekten Auswahl und Bestimmung der Stichprobe. Dumas und Redish geben hierzu folgenden Hinweis: „The people who come to test the product must be members of the group of people who now use or will use the product“. (1994:23) Demnach galt die Einteilung von Zielgruppen der Blaupunkt GmbH als erste Orientierung. (s. *Abb. 4.10*)



**Abbildung 4.10: Zielgruppeneinteilung der Blaupunkt GmbH**

Diese Zielgruppendefinition basiert auf folgenden Unterscheidungsmerkmalen: Anspruch an das Produkt, Art des Pkws, Einkommen, Bildungsstand, Alter und Ausgabebereitschaft. Die Teilnehmer des Benutzertests sollten sich an diesen Kategorien orientieren und möglichst gleichmäßig verteilen, um die Ergebnisse der Untersuchung in Zusammenhang mit den Eigenschaften der Probanden stellen zu können. Allerdings galt es nach einer Einschätzung der verfügbaren Freiwilligen für die Untersuchung einzelne Unterscheidungsmerkmale herauszunehmen und Untergruppen zusammenzustellen.

Um bereits in der Planung des Benutzertests mit diesen Variablen kalkulieren zu können, war es von Bedeutung, diese Eigenschaften der Freiwilligen im Vorfeld zu erfassen. Das so genannte *Screening* (vgl. DUMAS / REDISH 1994: 142) erfolgte auf einer Internetseite, die zum einen über die Studie und den Ablauf des Benutzertests informierte und zum anderen einen Fragebogen zur Anmeldung beinhaltete. Das in PHP programmierte Formular (s. Abb. 4.11) fragte sowohl die oben genannten biografischen Merkmal als auch Erfahrungswerte mit und Ansprüche an technische Geräte ab.

**Benutzertest "Innovative Fahrerinformationssysteme"**

Im Rahmen einer Masterarbeit der Universität Hildesheim sucht die Blaupunkt GmbH im Februar freiwillige ProbandInnen für eine Studie zur Akzeptanz von sprachbedienbaren Fahrerinformationssystemen.

Wir setzen auf Ihre *Neugier* auf Zukunftstechnologien und auch auf Ihre Hilfsbereitschaft, um uns bei dieser Studie zu unterstützen. Der Test kann in den Räumen der Universität Hildesheim, aber auch problemlos bei Ihnen vor Ort durchgeführt werden.

Vorab möchten wir Sie bitten, diesen Fragebogen auszufüllen. Er hilft uns den Test genauer zu planen und mit Ihnen einen Termin zu vereinbaren.

Sollten Sie noch Fragen haben, kontaktieren Sie bitte: Matthias Görtz / mgoe0088@rz.uni-hildesheim.de / 0177-2982149

**Persönliche Informationen**

Alter: ☐ 20 - 29 ☐ 30 - 39 ☐ 40 - 49 ☐ 50 - 59 ☐ 60 +

Geschlecht: ☐ weiblich ☐ männlich

PKW (Besitzen Sie kein Auto, so lassen Sie diese Zeile frei): ☐ Kleinwagen (Polo/Fiesta/Corsa) ☐ Untere Mittelklasse (Golf/Focus/Astra) ☐ Mittelklasse (Passat/A4/C-Klasse) ☐ Obere Mittelklasse (5er/Adi/E-Klasse) ☐ Oberklasse (7er/ABT/S-Klasse)

**Abbildung 4.11: Internetseite des Probanden-Screening ([www.neugier-innovation.de](http://www.neugier-innovation.de))**

In Emails und Aushängen wurde zur Teilnahme an dem Benutzertest aufgerufen und auf die Internetseite [www.neugier-innovation.de](http://www.neugier-innovation.de) hingewiesen. Das durch diese Seite unmittelbare Sammeln von Teilnehmern und ihren Zugehörigkeiten zu den unterschiedlichen Zielgruppen machte es möglich, die Auswahl der Teilnehmer gezielt zu steuern. Die Rekrutierung stieß auf große Rückmeldung und ergab geeignete Unterscheidungsmerkmale. So war die Gesamtheit der Teilnehmer in Gruppen bezüglich Ihres Alters, Geschlechts und Ihrer Erfahrung mit technischen Geräten einteilbar.

Über die optimale Anzahl von Teilnehmern gibt es in der Fachliteratur widersprüchliche Aussagen. Jakob Nielsen stellte die Faustregel auf, dass fünf Testteilnehmer ausreichen, um aussagekräftige Erkenntnisse bezüglich der Usability eines Systems zu treffen. (vgl. NIELSEN 1994: 393) Werner Schweibenz und Frank Thissen gehen davon aus, „dass etwa zehn bis zwölf Testteilnehmer einen sehr hohen Anteil der vorhandenen Usability-Mängel entdecken“ (SCHWEIBENZ / THISSEN 2002: 133). Zum einen lässt sich feststellen, dass in dieser Debatte keine eindeutige Zahl festgelegt wird. Zum anderen gilt dieser Benutzertest ja nicht direkt

der Aufdeckung von Usability-Mängeln. Vielmehr gilt es, die in der Experten-Evaluation entdeckten kritischen Elemente zu nutzen, um eine allgemeine Akzeptanz der Benutzer gegenüber einem System zu ermitteln. Insofern kann die Teilnehmerzahl wohl nie groß genug sein. So richtet sich die Anzahl der Teilnehmer dieses Tests primär nach der Durchführbarkeit im vorgegebenen Rahmen.

Insgesamt konnten 31 Probanden für eine Teilnahme an dem Benutzertest gewonnen werden. Hinsichtlich der oben genannten Merkmale ließ sich dabei folgende Einteilung in Untergruppen vornehmen:

		Geschlecht		Alter			technische Erfahrung	
		Männlich	Weiblich	20-29	30-39	40-59	unter ø	über ø
Geschlecht	Männlich	16	x	8	4	4	4	12
	Weiblich	x	15	8	3	4	9	6
Alter	20-29	8	8	16	x	x	4	12
	30-39	4	3	x	7	x	2	5
	40-59	4	4	x	x	8	7	1
technische Erfahrung	unter ø	4	9	4	2	7	13	x
	über ø	12	6	12	5	1	x	18

**Tabelle 4.2: Aufschlüsselung der Probandenauswahl**

Da in der oben aufgeführten Zielgruppeneinteilung der Blaupunkt GmbH auch der Bildungsstand als ein Unterscheidungsmerkmal gilt, ist bei der vorliegenden Probandenauswahl kritisch anzumerken, dass gerade die Altersgruppe 20-29 ausnahmslos aus Studenten besteht. Aussagen über die Population der unter 30-Jährigen lassen sich somit nur auf diese Untergruppe des Bildungsstands ableiten. Die im Vorfeld erfragte Erfahrung mit technischen Geräten schlüsselt sich in drei Kategorien auf. Zum ersten allgemeine Kenntnisse über Geräte wie CD-Spieler, MP3-Player oder Computer. Dann wurden die Teilnehmer nach ihrer Vorerfahrung mit Fahrerinformationssystemen oder den einzelnen Komponenten (Navigationssystem, Autotelefon, CD-Wechsler und Radio) gefragt. Als letzte Untergruppe wurde die bisherige Erfahrung mit Sprachdialogsystemen (wie z.B. auch automatischen Telefonauskünften) erhoben. Die Bewertung erfolgte jeweils auf einer Fünfer-Skala und ergab einen durchschnittlichen Gesamtwert von 8,7 (auf der Skala von 0 *keine* - 15 *sehr viel*). Die Verteilung dieses Werts über Altersgruppen und Geschlechter hinweg sind *Tabelle 4.2* zu entnehmen. Die Angaben zum Anspruch an das Produkt

ergaben keine Verteilung, die eine Merkmalsunterscheidung als möglich erschienen ließ. 13 Teilnehmer besitzen kein Auto, 12 einen Kleinwagen oder Wagen der unteren Mittelklasse und 6 Probanden fahren in einer höheren Fahrzeugklasse. Auch diese Kategorie stellte sich aber im weiteren Verlauf der Untersuchung nicht als differenzierend dar. Fragen zum Einkommen wurden auf der Screening-Internetseite nicht erhoben, um eine gewisse Diskretion im Vorfeld des Benutzertests zu wahren.

Die Auswahl der Probanden erfolgte auf Basis dieser im Vorfeld erfassten Merkmale. Dabei war allen Teilnehmern gemein, dass sie in keinem direkten Kontakt zu den Unternehmen Blaupunkt bzw. Charamel standen. Die Registrierung für den Benutzertest signalisiert aber ein generelles Interesse an den Produkten der Blaupunkt GmbH und Fahrerinformationssystemen im Allgemeinen. Mit der erwähnten Einschränkung des Bildungsstands, kann man insofern durch die gewonnene Probendenauswahl die Ergebnisse des Benutzertests auf unterscheidbare Untergruppen bzw. Zielgruppen anwenden.

#### **4.4.2. Aufbau und Durchführung**

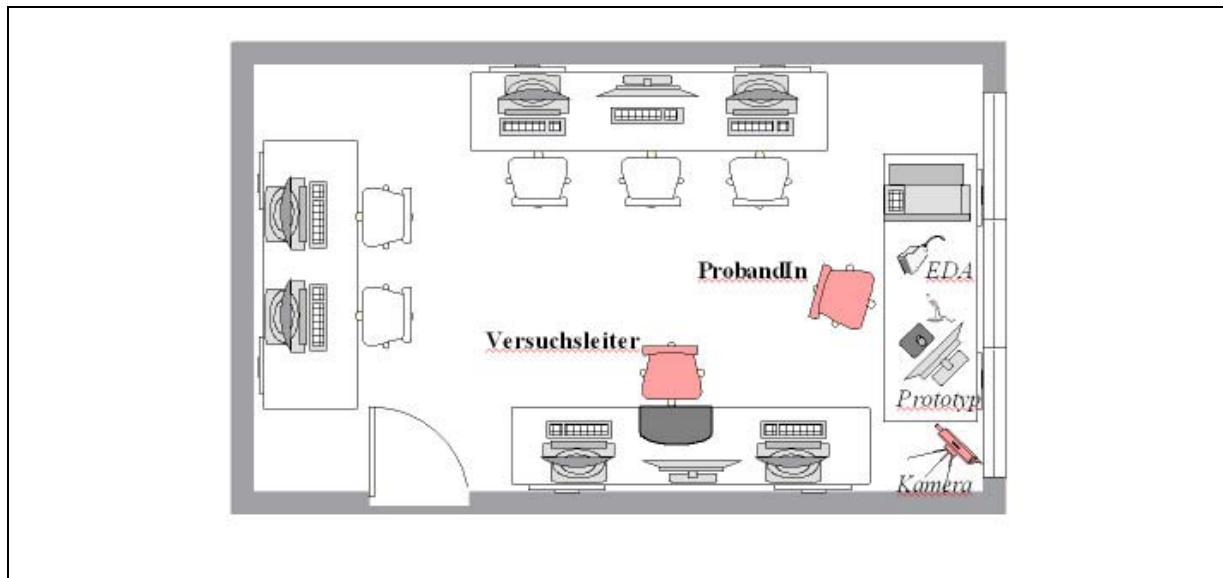
Bei der Konzeption eines Benutzertests spielen die gewählte Methodik und der Versuchsaufbau eine entscheidende Rolle. Eventuelle Fehlentscheidungen in ihrer Konzeption können die Ergebnisse des Tests verfälschen und somit ihre Aussagekraft schmälern. Für die Untersuchung der Hypothesen dieser Arbeit war es zunächst von Bedeutung, die Vergleichbarkeit der Beobachtungen und subjektiven Bewertungen der Probanden zu garantieren. Die Anpassung der Prototypen und die differenzierte Konzeption der kontrollierten Aufgabenszenarien sollen diese Untersuchungsbedingungen garantieren, die besonders für die Arbeitshypothesen 1 – 3 von großer Bedeutung sind. Zusätzlich spielt die Berücksichtigung des Anwendungskontexts der untersuchten Prototypen eine entscheidende Rolle. Die heuristische Evaluierung verwies diesbezüglich auf spezifische Richtlinien zur Vermeidung von Ablenkung im Fahrzeug. Dies ließ für den Versuchsaufbau die Frage aufkommen, ob der Usability-Test in einem Fahrsimulator oder gar einem Testwagen durchgeführt werden sollte, um zu berücksichtigen, inwiefern die Akzeptanz eines

virtuellen Charakters im tatsächlichen Fahrerumfeld vorhanden ist. Gegen einen Versuchsaufbau im Fahrsimulator sprach jedoch, dass die heute verfügbaren Anzeigeflächen in der Mittelkonsole sich bis zu einem wahrscheinlichen Einsatz der untersuchten Technologie grundlegend verändern werden. Gerade in den Fahrzeugen der höheren Mittelklasse und Oberklasse zeichnet sich eine Veränderung der Anzeigeinstrumente im Kombi-Instrument ab. Dort, wo heute noch analoge Zeiger den Tankstand oder die Geschwindigkeit anzeigen, werden schon bald flexible, digitale Displays eingesetzt, die situationsspezifisch ihren Inhalt anzeigen können. Ein Versuchsaufbau, der die heutigen Displayflächen verwendet, würde also ein Anwendungsszenario simulieren, das bezüglich des Orts und der Art der Visualisierung stark einschränkt. Somit wurde ein Szenario gewählt, das sich allein auf das Dialogsystem konzentriert und eine Bedienung im stehenden Fahrzeug simuliert. Die Vorgehensweise in Form von geführten Aufgabenszenarien sollte dafür sorgen, dass sich die Probanden unabhängig von heutigen Displayflächen vorstellten, in einem Auto mit diesen Systemen zu arbeiten.

Der Versuch fand demnach in einem Usability-Labor statt. „Tests, die in Usability Labs durchgeführt werden, lassen sich besonders genau auswerten. Dennoch hat auch das Testen in Usability Labs Nachteile.“ (PREIM 1999: 252) Eine gewisse Distanz zum gewohnten Anwendungsumfeld und eventuelle Einflüsse durch den sog. *Hawthorne-Effekt*<sup>15</sup> sind mögliche negative Aspekte der Laborumgebung, aber die oben genannten Vorteile und die Tatsache, dass Probanden einen geringeren Anreiseaufwand haben und eine Fokussierung auf die beiden Prototypen und die unterschiedlichen Visualisierungsformen möglich ist, überwiegen. Deshalb wurden die Probanden bei ihrer Aufgabenbearbeitung mit den beiden Prototypen in einem fest installierten Usability-Labor (s. Abb. 4.12) beobachtet.

---

<sup>15</sup> Menschen neigen dazu, sich in einer Versuchssituation engagierter und somit anders zu verhalten, als wenn sie etwas unbeobachtet tun. (vgl. PREIM 1999: 196)



**Abbildung 4.12: Raumaufteilung des Usability-Labors**

Das Labor ist dabei so aufgeteilt, dass an einem Arbeitsplatz der Proband mit dem jeweiligen System interagiert, während eine Videokamera den Versuchsaufbau aufzeichnet und der Versuchsleiter im Hintergrund die Ereignisse dokumentiert und die Aufgabenszenarien beschreibt.

Nach einer Begrüßung und Einweisung in die allgemeine Thematik der Untersuchung wurde der Proband in den Versuchsraum begleitet und über die Vorgehensweise, die Datenaufzeichnung und –auswertung informiert. Jeder Proband erklärte daraufhin schriftlich zunächst seine Einwilligung zur Verwendung der persönlichen Daten. Der Versuchsleiter zeigte anschließend einige Beispiele der möglichen Implementierungen von FIS und ihrer Darstellbarkeit auf Displayflächen im Fahrzeug. Nachdem dem Teilnehmer die Anwendungssituation klar war, wurde er gebeten sich an den Arbeitsplatz zu begeben. (s. Abb. 4.13)





**Abbildung 4.13: Versuchsaufbau mit Proband**

Dort wurden ihm die Geräte und die Bedienung des Prototyps erläutert. Der Prototyp war auf einem Laptop mit 15“ Display installiert und mit Lautsprechern und Mikrofon versehen. Die PTT-Taste zur Initialisierung einer Eingabe wurde durch den linken Mausknopf auf der Laptop-Tastatur aktiviert. Eine Videokamera schräg hinter dem Laptop zeichnete den Benutzer und seine Reaktionen während des Tests auf. Zusätzlich wurde der Proband an ein Gerät zur Messung physiologischer Daten angeschlossen.

#### **Exkurs: Messmethoden psychophysiologischer Daten**

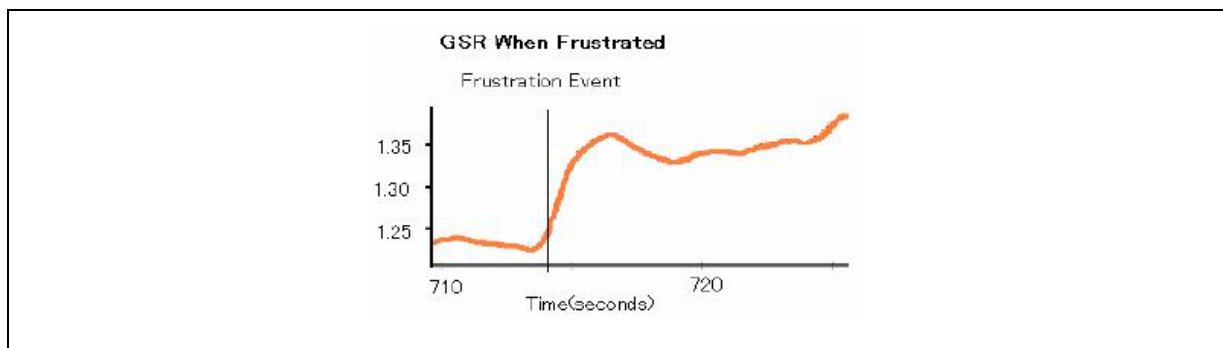
Die Arbeitshypothesen 2 und 3 machen deutlich, dass die empfundene Attraktivität eines Systems und die Frustration des Benutzers bei der Aufgabenerledigung entscheidende Faktoren dieser Untersuchung darstellen. Will man der Vermutung auf den Grund gehen, dass Spaß und ästhetisches Empfinden ein entscheidendes Merkmal der Usability eines Systems sind, muss man Messmethoden verwenden, die dies prüfen können. Der Post-Task-Fragebogen zur Erhebung subjektiver Werte ist dabei eine weit verbreitete Methode. Allerdings orientieren Benutzer „ihre Aussagen häufig an dem, was sie meinen, was man von ihnen hören will oder was sozial akzeptabel

erscheint.“ (SCHWEIBENZ / THISSEN 2002: 156) Die reine Auswertung der Bewertung der Probanden kann demnach teilweise erwünschte oder generell falsche Ergebnisse liefern.

Um die Aussagen der Probanden zu verifizieren bzw. objektiven Daten gegenüberzustellen, müssen messbare Kennzahlen entwickelt werden, die eine Einschätzung der empfundenen Gemütslage des Teilnehmers wiedergeben. Ein mögliches Kriterium z.B. bei der Videoanalyse ist die Körpersprache bzw. Mimik des Probanden. So genannte kritische Ereignisse können von einem Evaluator beobachtet, klassifiziert und später interpretiert werden. (vgl. HASSENZAHN / SEEWALD 2004: 142) Allerdings findet hier eine subjektive Analyse durch den Auswertenden statt, was objektive Quantität erschwert. Ein anderes Instrument bietet die Psychophysiologie in Form von Messung entsprechender Kennwerte wie elektrodermale Aktivität und Herzfrequenz. Lin et al. sehen in der Verwendung dieser Messwerte eine neue Methode der Usability-Evaluierung, „based on three kinds of data: user-subjectivity, task performance data and physiological data, which not only evaluates the system's effectiveness and efficiency, but also takes physiological reactions into account.“ (2005: 7) Psychophysiologische Emotionsstudien versuchen also das Bedürfnis nach Objektivität bei der Erfassung der Befindlichkeit des Menschen zu bedienen. Boucsein legt für die Messung von Emotionen beim Menschen drei Beobachtungs- bzw. Messebenen zugrunde: die subjektive Komponente, die Verhaltenskomponente (äußerlich beobachtbare motorische Veränderungen) und die Neuro-endokrine Komponente (somatische Veränderung). (vgl. 1988: 311) Ein häufig evaluierter Kennwert dieser dritten Ebene ist die elektrodermale Aktivität (EDA). Die EDA bezeichnet die elektrischen Phänomene der Haut und umfasst sowohl die aktive als auch die passive Leitfähigkeit. Diese Hautleitfähigkeit lässt sich in tonische (*level*) und phasische (*response*) Anteile unterscheiden und basiert im Wesentlichen auf epidermalen und Schweißdrüsenmechanismen. (vgl. a.a.O.: 2ff.) Die leichte Messbarkeit führte zu einer gewissen Popularität dieser Kennwerte, die gerade in den letzten Jahren auch häufiger in der Usability-Forschung Einzug erhielten.

Dabei ist allerdings noch keine einheitliche Emotions-Taxonomie oder eindeutige Interpretationsweise der Daten erreicht worden. Lin et al. formulieren in ihrer Untersuchung die Frage: „Do physiological data relate to traditional usability indexes?“ und stellen die physiologischen Messwerte den klassischen Usability-Kennzahlen der beobachteten Leistungsdaten und subjektiven Bewertung der Probanden gegenüber. Sie kommen zu dem Schluss, dass die Hautleitfähigkeit ein direkter Indikator für Frustrationsereignisse darstellt und ein geeignetes Messinstrument für die Beanspruchung der Benutzer während eines Usability-Tests ist. Allerdings bleibt weiterhin eine Standardisierung dieser Methodik offen. (vgl. Lin et al. 2005) Dennoch stellen gerade diese Hinweise auf Messbarkeit von Nutzerfrustration und emotionaler Aktivierung einen interessanten Ansatz für die Prüfung der Hypothesen dieser Arbeit dar.

Lin et al. beschreiben in ihrer Studie (s.o.) einige charakteristische EDA-Kurven, in denen Frustrationsereignisse erkennbar sind.



**Abbildung 4.14: Phasische EDA bei Frustration des Probanden**

(LIN et al. 2005: 7)

Der Ausschlag der gemessenen Hautleitfähigkeit (s. Abb. 4.14) wird von den Autoren als Zeichen für die Frustration des Benutzers gewertet. Geht man von dieser ersten Überlegung zur phasischen EDA (*Response*) aus, so lässt sich für die Studie schlussfolgern: Je mehr solcher phasischen Reaktionen der Proband zeigt, desto häufiger geriet er in kritische Situationen und reagierte er gereizt auf das Dialogsystem. Zusätzlich formulierte schon Boucsein die Bedeutung tonischer EDA-Maße. In seinen Untersuchungen stellten sich diese

Werte als „sensible und valide Indikatoren mentaler, möglicherweise sogar spezifischer emotionaler Beanspruchung“ heraus. (1988: 421) Der tonische EDA-Wert beschreibt das Niveau, d.h. keine sequentiellen Reaktionen, sondern die Entwicklung des Messlevels für den Zeitraum der Betrachtung. Die erhobenen Daten des Benutzertests können auf die Entwicklung dieses tonischen Werts (*Level*) untersucht werden, wobei ein starker Anstieg dieses Kennwerts mit hoher emotionaler Beanspruchung gleichzusetzen ist. Die beiden EDA-Werte sollen anhand der in dieser Untersuchung gewonnenen Messwerte exemplarisch mit den oben präsentierten Leistungsdaten und subjektiven Bewertungen des Fragebogens gegenübergestellt werden. Beispielhaft kann dadurch analysiert werden, inwiefern die genannten Werte aussagekräftig sind und eine Tendenz zu erkennen ist, die Aufschluss darüber gibt, ob die gemessene Emotionalität mit den Erkenntnissen der Beobachtung und den Aussagen der Probanden übereinstimmen.

Um mit der Erhebung quantitativer Daten eine Verifizierung der subjektiven Bewertung der Gemütszustände der Probanden zu erhalten, wurde in diesem Usability-Test also ein Messgerät physiologischer Daten verwendet. Dabei war es von Bedeutung, einen komplizierten und aufwändigen Versuchsaufbau zu umgehen. Vor diesem Hintergrund wurde eine Zentraleinheit für Biofeedbackmessungen des *Wild Divine Project*<sup>16</sup> in Kombination mit einer *Grapher*<sup>TM</sup>-Software verwendet. Diese Elemente eines Produkts aus der Biofeedback- bzw. Unterhaltungsbranche bietet käuflich erhältliche Sensorik (s. *Abb. 4.15*), die es erlaubt per USB-Anschluss an einem gewöhnlichen PC die *Galvanic Skin Response* (EDA) und *Heart Rate* (Beats Per Minute) zu messen.

---

<sup>16</sup> Das Computerspiel *The Journey to Wild Divine* verwendet Biofeedback-Sensoren, um über die physiologischen Werte des Benutzers das Spiel zu steuern. (vgl. URL: The Wild Divine Project)



**Abbildung 4.15: Biofeedbacksensoren zur Messung physiologischer Daten**

Der Proband wird während der Bearbeitung der Testaufgaben mit den drei mittleren Fingern der linken Hand an das Messgerät angeschlossen. Über den Mittelfinger wird dabei die Herzfrequenz, über Zeige- und Ringfinger die Hautleitfähigkeit abgenommen. Die Grapher™-Software zeichnet diese Werte auf einer Zeitskala auf. Die Herzfrequenz wird in diesem Graphen in Herzschlägen pro Minute (BPM) und die Hautleitfähigkeit in  $\mu\text{Mho}$  (*Units of Measure*) angegeben.

Den Beginn einer jeden Untersuchung markierte die *Einleitung*, in der durch den jeweiligen Prototypen eine Erklärung der Funktionsweise vorgelesen wurde. Diese Eingewöhnungsphase wurde genutzt, um eine Baseline<sup>17</sup> der physiologischen Daten 90 Sekunden lang aufzuzeichnen. Daraufhin begann der Versuchsleiter, von seinem Platz aus dem Teilnehmer eine Geschichte zu erzählen, die dazu diente, ihn in die fiktive Anwendungssituation zu versetzen. Innerhalb dieses Szenarios wurde dann der Benutzer gebeten, die verschiedenen Aufgaben zu lösen. In einer Voruntersuchung zum Versuchsaufbau hatte sich ergeben, dass Aufgabenzettel den Probanden zum Lesen anregen und den Blick von der Anwendung ablenken. Diese Methode ermöglicht es dem Benutzer, sich das Anwendungsszenario vorzustellen und den Anweisungen des Versuchsleiters zu folgen, ohne dabei zusätzliche Arbeitsschritte ablesen zu müssen. Während der Proband die Aufgaben bearbeitete, wurden Anzahl der Eingaben, der Erkennungsfehler und Art der Interaktionsschritte

---

<sup>17</sup> Diese „Normalkurve“ soll einen Eindruck vermitteln, wie der Kurvenverlauf ohne größere Beanspruchung bei dem jeweiligen Probanden aussieht.

beobachtet. Eine direkte Interaktion zwischen Proband und Versuchsleiter wurde vermieden. War der Teilnehmer nicht in der Lage einen Aufgabenschritt zu erfüllen, so wurde zum nächsten Punkt übergeleitet und im Notfall helfend eingeschritten.

Nach Beendigung des ersten Aufgabenszenarios wurde der Proband gebeten, in einem Fragebogen (s. CD-ROM) seine Eindrücke bezüglich der Interaktion mit dem Prototypen und der Darstellung der *HelpForm*, also der Befehlsliste, wiederzugeben. Diese Befragung war in folgende Aspekte unterteilt:

*Item 1 - Usability:* Usability des Prototypen

*Item 2 - Emotion:* Empfindungen während der Benutzung

*Item 3 – Overall:* Gesamtbewertung des Systems

*Item 4 – Difficulty:* Schwierigkeiten bei der Aufgabenbewältigung und Hilfe

*Item 5 – Buy:* Kauf- und Ausgabebereitschaft

Die ersten beiden Items setzen sich aus sechs bzw. sieben Einzelfragen auf einer Siebener-Skala (--- / 0 / +++) zusammen. Zur Bewertung des Systems wurde die Systematik des ISONORM-Fragebogens übernommen. Einer negativen Aussage steht das positive Gegenteil am anderen Ende der Skala gegenüber. Je nachdem zu welcher Aussage der Proband eher tendiert, kann er sein Urteil über das System bezüglich dieser Kategorie fällen.

Das System ...	---	--	-	0	+	++	+++	Das System ...
ist schlecht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ist gut.

Beispiel 1

Bei dem angegebenen Beispiel beurteilt der Befragte das System zwar als gut, sieht jedoch noch Verbesserungsmöglichkeiten. (vgl. URL: ERGO-ONLINE)

Das erste so behandelte Item ist eine Essenz aus dem Fragebogen, der in der Experten-Evaluierung bereits eingesetzt wurde. Es soll anhand von Zustimmung oder

Ablehnung von Aussagen den Eindruck des Benutzers wiedergeben, inwiefern er das System gebrauchstauglich fand. Das zweite Item erfasst, wie sich der Proband während der Benutzung des Systems fühlte. Das hier verwendete Vokabular lehnt sich an das Berliner-Alltagssprachliche-Stimmungs-Inventar (BASTI) (vgl. SCHIMMACK 1997) an. Item 3 bietet die Gelegenheit, neben dem Gesamteindruck auf der Skala (*Wie hat Ihnen das System insgesamt gefallen?: gar nicht – sehr gut*) in offenen Fragen besonderes Gefallen bzw. Missfallen auszudrücken. Die anschließenden Fragen beziehen sich auf Aufgabenteile, die besonders schwer gefallen sind und darauf, ob das System bei der Lösung eventueller Schwierigkeiten behilflich war. Das letzte Item betrachtet die Kaufbereitschaft des Probanden und wie viel Geld er bereit wäre, für ein solches System auszugeben.

Nach dieser umfassenden Bewertung des ersten Prototypen und einer kleinen Pause wurde dann das zweite System gestartet und der Proband zunächst wieder durch ein Aufgabenszenario geführt. Die Geschichte wurde fortgeführt mit dem Hinweis, dass sich der Proband nun in einem anderen Fahrzeug mit einem von der Gestaltung her abweichenden Fahrerinformationssystem befindet. Auch dieses Dialogsystem wurde nach Abschluss der Aufgabenphase in einem Fragebogen bewertet. Dieser zweite Komplex weicht nur in einem Aspekt von dem ersten ab. Er beinhaltet zusätzlich *Item 6 – Preference*. Der Benutzer kann hierbei angeben, welches System er bevorzugt in seinem Fahrzeug einsetzen würde und dies begründen.

Der gesamte Benutzertest dauerte somit ca. 60 Minuten. Abschließend wurde den Probanden in einem freien Gespräch mit dem Versuchsleiter die Gelegenheit gegeben, zusätzliche Eindrücke wiederzugeben und zu diskutieren. 68% (21) der Probanden wurden in dem oben beschriebenen Labor beobachtet. Die Mobilität der eingesetzten Geräte ermöglichte zudem, in jedem beliebigen Büroraum den Versuchsaufbau nachzubilden. So wurden 32% (10) der Probanden in privaten oder anderen öffentlichen Räumlichkeiten bei ähnlichem Aufbau beobachtet. Gerade diese Flexibilität des Aufbaus erwies sich insofern als vorteilhaft, als auch Probanden erreicht werden konnten, denen eine Anreise zum Usability-Labor nicht möglich war. In einem Zeitraum von sechs Wochen konnten die Daten und Meinungen von 31 Probanden eingeholt werden. Die Ergebnisse werden im *Abschnitt 4.4.3.* dargelegt.

### 4.4.3. Ergebnisse

Die Erhebung der Daten des Usability-Tests gliedert sich in drei Dimensionen. Ein wesentliches Merkmal zur Messung der Usability ist die Art und Weise, wie der Benutzer mit den Systemen interagieren und die ihm gestellten Aufgaben bearbeiten konnte. Diese Leistungsdaten (*Task Performance Data*) geben einen quantitativen Einblick in die tatsächliche Effektivität und Effizienz eines Prototyps. Der oben vorgestellte Post-Task-Fragebogen ermöglicht eine subjektive Bewertung der Usability und somit die Erhebung der Akzeptanz und vor allem der Präferenz der untersuchten Prototypen. Zusätzlich sollen versuchsweise die physiologischen Daten der Aufgabebearbeitungsphase herangezogen werden, um eine Messbarkeit der Emotionalität zu prüfen.

#### 4.4.3.1. Leistungsdaten (Task Performance Data)

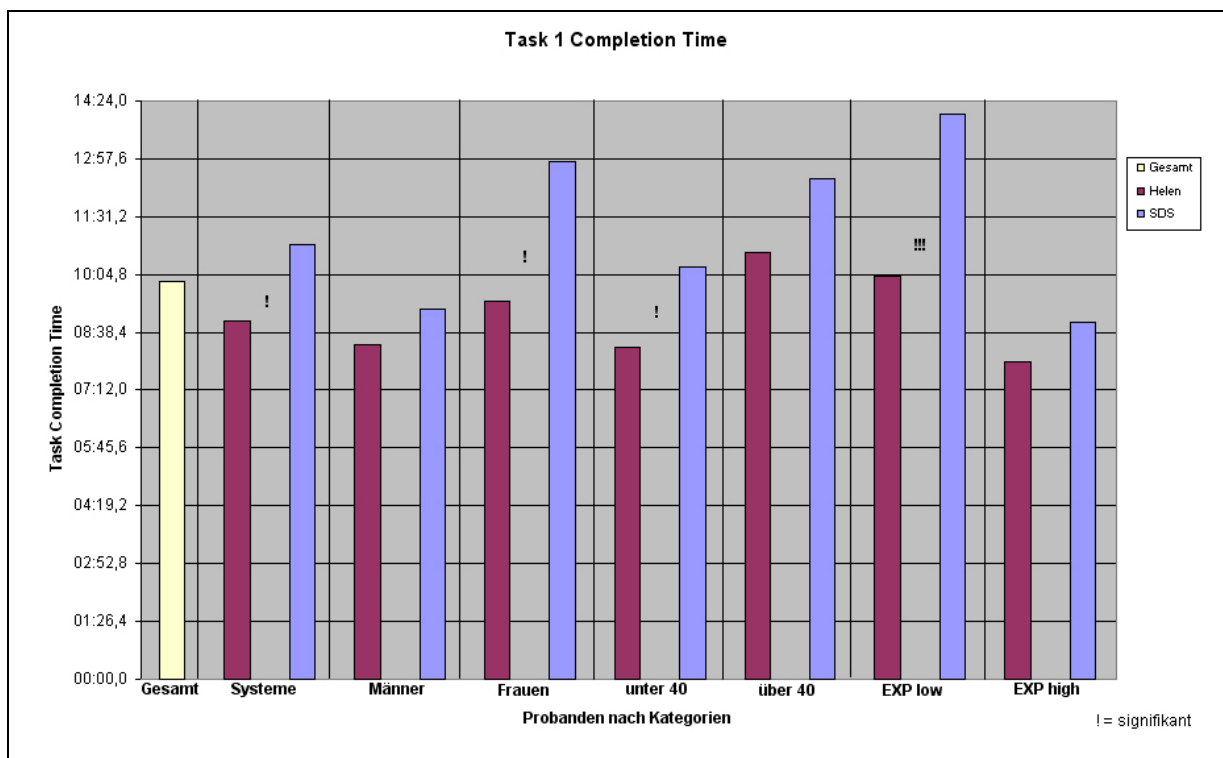
Die erhobenen Leistungsdaten entstehen aus den Aufzeichnungen des Versuchsleiters während des Tests und der Auswertung der Videoaufzeichnungen. Für die Messung der Effektivität und Effizienz der Benutzung können bezogen auf die Aufgabenszenarien der Untersuchung folgende Kennwerte erhoben werden: Zeit, die benötigt wird um eine Aufgabe zu erledigen (*Task Completion Time*), Anteil der erfolgreich bearbeiteten Aufgaben (*Task Completion Rate*), die Anzahl der Bedienfehler (Task Errors) und die Spracherkennungsrate (*Recognition Rate*) der Systeme während der Aufgabebearbeitung. Diese Daten geben Aufschluss darüber, wie gut ein Proband im Vergleich zu anderen Teilnehmern, aber auch im Vergleich zum anderen System, mit einem Prototyp zu Recht kam.

Das erste Aufgabenszenario wurde von 48% der Probanden (15) mit dem Prototyp *Helen* bearbeitet, während 52% (16) die gleiche Aufgabenstellung im ersten Versuchsteil mit dem *SDS Demonstrator* arbeitete. Somit ist dafür gesorgt, dass eventuelle Lerneffekte und Änderungen in der Wahrnehmung vom Erstkontakt mit einem System zur zweiten Versuchsphase mit dem anderen Prototyp bei beiden Systemen gleichmäßig auftreten. Die generell in der zweiten Aufgabe verbesserte *Task Performance* ist somit bezogen auf die Bewertung beider Prototypen



ausgeglichen. Auch innerhalb der Merkmalsunterscheidungen (Geschlecht, Alter, Technische Erfahrung (EXP)) der Probanden war jeweils die Verteilung auf die Systeme ausgeglichen, um eine größtmögliche Vergleichbarkeit sicherzustellen.

Aufgabenstellung 1 dauerte im Durchschnitt aller Probanden 9:53 Minuten (Standardabweichung 2:44). *Abbildung 4.16* zeigt neben diesem Mittelwert, die jeweilige Gegenüberstellung der Untergruppen der Stichprobe.<sup>18</sup> Dabei ist zwischen systemspezifischen und kategorischen Faktoren zu unterscheiden, d.h. den Ergebnissen der unterschiedlichen Systeme und denen der Merkmalsunterschiede der Probanden.



**Abbildung 4.16: Überblick Task 1 Completion Time**

Die Grafik zeigt (von links nach rechts), dass die Probanden, die in dieser Aufgabe mit dem System *Helen* gearbeitet haben, 8:54 Minuten benötigten, während diejenigen mit dem *SDS Demonstrator* 10:49 Minuten und somit signifikant ( $p=0,026$ ) länger brauchten. Männliche Probanden beendeten das erste Aufgabenszenario

<sup>18</sup> Die Ausrufezeichen signalisieren die Ergebnisse der statistischen Analyse *t-Test zweier Stichproben mit unterschiedlicher Varianz* (Excel).

! = signifikant ( $p \leq 0,05$ ) - !! = sehr signifikant ( $p \leq 0,01$ ) - !!! = höchst signifikant ( $p \leq 0,001$ )

durchschnittlich nach 8:43 Minuten, die Probandinnen nach 11:02 Minuten. Dabei waren jedoch die Leistungsdaten in Bezug auf den verwendeten Prototyp von entscheidender Bedeutung. So ist bei den männlichen Teilnehmern zwar auch zu beobachten, dass mit dem System *Helen* knapp eine Minute schneller gearbeitet wurde, diese Differenz ist jedoch nicht signifikant. Bei den Frauen hingegen ist eine deutliche Korrelation von Bearbeitungsdauer und verwendetem System zu erkennen. Die Erledigung der Aufgaben mit System *SDS* (12:53 Minuten) benötigte signifikant mehr Zeit als mit Prototyp *Helen* (9:25 Minuten /  $p=0,036$ ).

Auch bezüglich des Alters der Teilnehmer ergaben sich signifikant unterschiedliche Bearbeitungszeiten. Probanden unter 40 Jahren benötigten durchschnittlich 9:19 Minuten, die Altersgruppe 40 – 60 hingegen 11:33 Minuten. Bei den Teilnehmern der höheren Altersgruppe ist keine Signifikanz messbar, da bei der geringen Größe der Stichprobe die Werte zu stark variierten. Die Testpersonen unter 40 Jahren arbeiteten mit dem anthropomorphen Dialogsystem zwei Minuten und somit signifikant ( $p=0,034$ ) schneller.

Diejenigen Probanden, die im Screening eine überdurchschnittliche Vorerfahrung angaben (EXP high), hatten nach 8:21 Minuten die Aufgabenstellung 1 erledigt und waren damit 3:25 Minuten schneller als Probanden mit wenig Erfahrung im Umgang mit technischen Geräten. Die Gruppe der technisch unerfahrenen Teilnehmer ist besonders gut mit System *Helen* zurechtgekommen. Hier lag die Differenz zwischen den beiden Prototypen bei über vier Minuten, was ein höchst signifikantes Ergebnis ist.

Die Bearbeitungsdauer allein lässt keine Schlüsse bezüglich der Bedienbarkeit eines Sprachdialogsystems zu. Erst die Angabe, ob die Aufgaben auch erfolgreich gelöst wurden, ergeben ein vollständiges Bild der Effektivität und Effizienz der Mensch-Maschine-Interaktion. *Abbildung 4.17* stellt bezüglich des Grads der Aufgabenerfüllung die oben genannten Untergruppen der Probanden einander gegenüber.

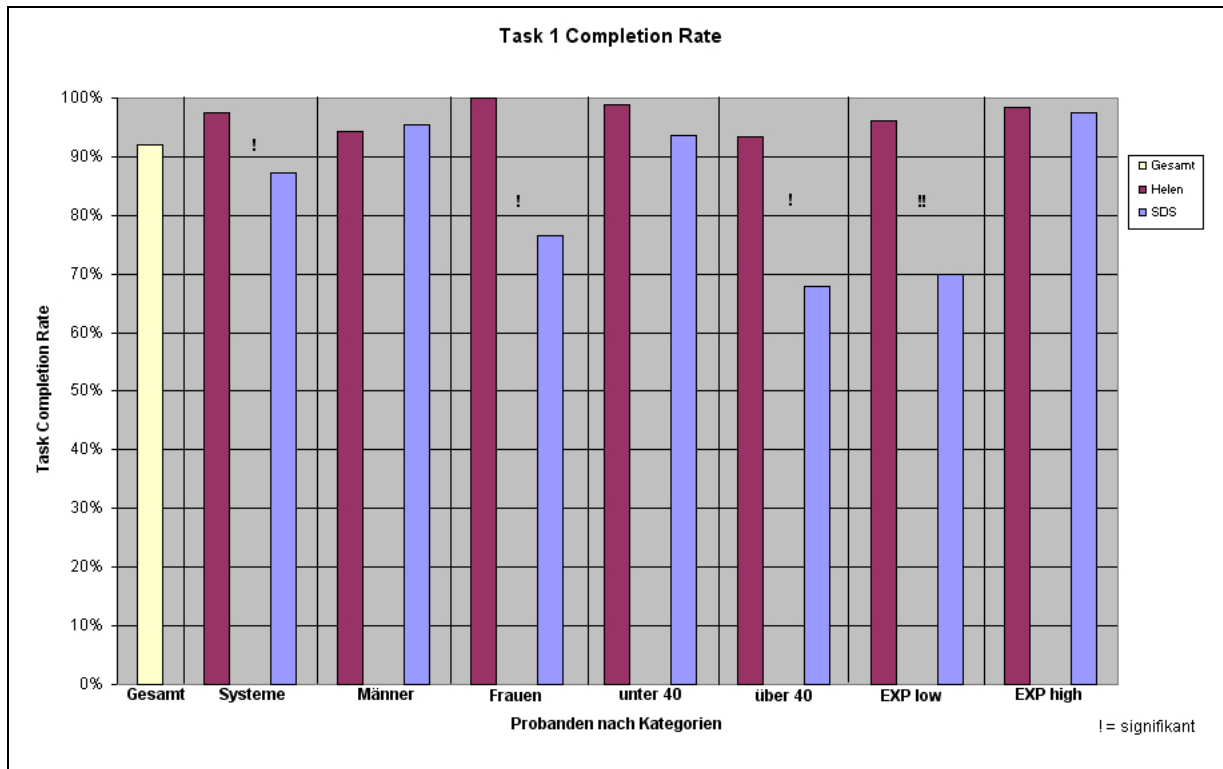


Abbildung 4.17: Überblick Task 1 Completion Rate

Die durchschnittliche *TCR* lag in der ersten Aufgabenphase bei 92 Prozent. Die Probanden, die das System *Helen* verwendeten, konnten dabei im Schnitt 97% der Aufgaben lösen, Testpersonen mit dem *SDS Demonstrator* nur 87% ( $p=0,043$ ). Dabei ist diese Leistungssteigerung der Arbeit mit dem anthropomorphen Prototypen unterschiedlich auf die Untergruppen der Probanden verteilt. Männliche Teilnehmer erzielten mit beiden Systemen beinahe gleiche Aufgabenerfüllungsraten, Frauen scheinen signifikant besser mit einem virtuellen Charakter zu arbeiten. 100% der Aufgaben lösten sie mit System *Helen*, lediglich 77% mit dem *SDS*-Prototypen ( $p=0,015$ ). Die über 40-jährigen Teilnehmer zeigten ähnliche Leistungswerte. Sie lösten sogar 25% mehr Aufgaben, wenn Sie mit dem anthropomorphen Interface-Agenten interagierten ( $p=0,05$ ). Die technische Vorerfahrung stellt erneut ein signifikantes Unterscheidungsmerkmal dar. Probanden, die wenig technisches Know-how mitbrachten, lösten in Interaktion mit *Helen* 96% der Aufgaben, während sie ohne einen virtuellen Ansprechpartner nur 68% lösen konnten. Diese Differenz ist mit einer Irrtumswahrscheinlichkeit von  $p=0,01$  sehr signifikant. Bei der Probandengruppe mit viel Vorerfahrung war hingegen kein Unterschied in der Leistung mit den beiden

Systemen zu beobachten. Betrachtet man die beiden oben dargestellten Ergebniskategorien, so ist in diesen ersten Datenreihen eine direkte Abhängigkeit zwischen den Werten Aufgabenbearbeitungsdauer und Erfüllungsrate zu beobachten. Ein kausaler Zusammenhang zwischen *TCR* und *TCT* war zu erwarten und stärkt die Aussagekraft der erhobenen Messwerte.

Nach der ersten Aufgabenbearbeitung mit System 1 und der anschließenden Bewertung wechselten die Probanden nun zu dem jeweils anderen Prototypen und durchliefen das zweite Testszenario.

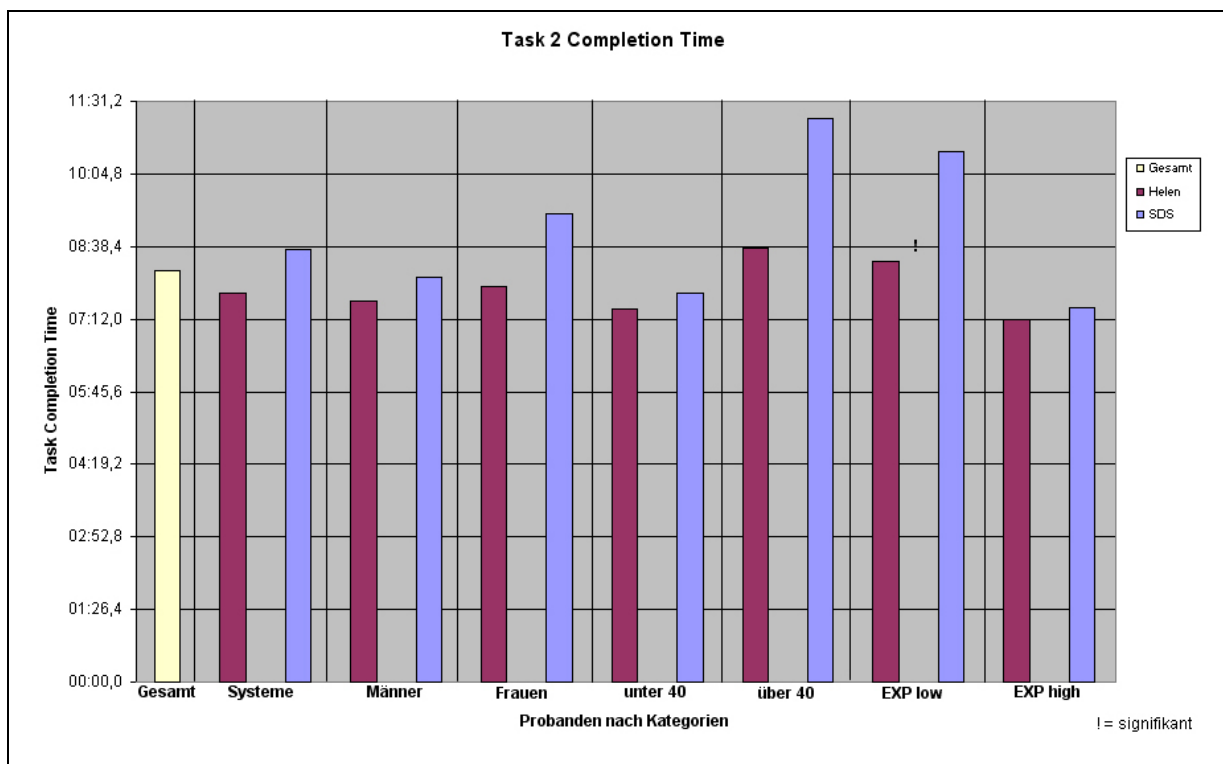


Abbildung 4.18: Überblick Task 2 Completion Time

Wie Abb. 4.18 zeigt benötigten die Probanden durchschnittlich 8:10 Minuten zur Bearbeitung der zweiten Aufgabenstellung (Standardabweichung 1:43), davon entfielen 7:43 Minuten auf den Prototypen *Helen* und 8:35 Minuten auf den *SDS Demonstrator*. Die Tendenzen der ersten Beobachtungen setzen sich fort. Jedoch verringert der Lerneffekt nicht nur die Arbeitszeit sondern auch systemspezifische Phänomene. Es ist zwar weiterhin die Tendenz vorhanden, dass mit *Helen* jeweils

schneller gearbeitet wird als mit System *SDS*, die gemessenen Unterschiede sind aber kaum signifikant. Lediglich die bisher auffälligste Untersuchungsgruppe weist auch hier eine signifikante Besserstellung des Prototypen *Helen* auf. Technisch weniger erfahrene Probanden benötigten nur 8:20 Minuten, um mit dem anthropomorphen Prototypen alle Aufgaben des zweiten Szenarios zu bearbeiten, während sie mit dem reinen Sprachdialogsystem 10:31 Minuten brauchten ( $p=0,028$ ). Die hohe durchschnittliche *Task Completion Rate* (97%) der zweiten Aufgabe ist ein Zeichen für die erfolgreichere Aufgabenbearbeitung der zweiten Runde. Bezüglich der Systeme und Untergruppen der Stichprobe waren aber kaum Unterschiede messbar. Die Erfolgsquote von System *Helen* lag mit 98% leicht höher als die mit Prototyp *SDS* bei 96%. Männer und Frauen erzielten beide eine *TCR* von 97%, wobei der anthropomorphe Prototyp erneut leicht besser abschnitt. In der Kategorie *EXP* unterschieden sich die Ergebnisse nur um 2%, die Altersklassen unter 40 Jahren und 40-60 Jahren um 3%. Weibliche Probandinnen lösten mit dem *SDS Demonstrator* 98% der Aufgaben, die dazu gehörige Vergleichsgruppe mit *Helen* 97%. Auch bei den Männern waren die 2% Unterschied der *TCR* im zweiten Aufgabenszenario zwischen den beiden angewandten Dialogsystemen im Gegensatz zum unmittelbaren Erstkontakt mit einem sprachbedienbaren Fahrerinformationssystem minimal.

Die anfangs genannten Kriterien der Spracherkennungsrate und Fehleranzahl zeigten keinerlei kategorien- bzw. systemspezifischen Ergebnisse. Die Erkennungsrate lag durchschnittlich bei 87%. In Anbetracht der Tatsache, dass es sich für die Probanden –abgesehen von einer kurzen Einführungsphase – um den direkten Erstkontakt mit den Sprachdialogsystemen handelte, ist dies ein Kennzeichen der hohen Qualität des Spracherkenners. Die Anzahl der Bedienfehler variierte in der Gesamtpopulation stark. Die Videoauswertung erfasste die folgenden Fehlerarten: *Time In (TI)*, *Time Out (TO)* und *Out Of Vocabulary (OOV)*. Damit sind jeweils Bedienschritte der Probanden gemeint, die insofern zu Erkennungsfehlern führten, als entweder zu früh (*TI*) bzw. zu spät (*TO*) gesprochen oder ein nicht zur Verfügung stehender Sprachbefehl verwendet wurde. Allgemein war zu beobachten, dass die Anzahl der Fehler vom ersten Aufgabenszenario zum zweiten deutlich abnahm. Die frustrierenden Fehlersituationen der ersten Versuche führten zu einem hohen

Lerneffekt. Die Gewöhnung an die Bediensystematik und an die Spracherkennung des Fahrerinformationssystems schlug sich in allen untersuchten Kategorien nieder. Die deutliche Verbesserung der Leistung aller Probanden führte dazu, dass die systemspezifischen Unterschiede zwar tendenziell erkennbar blieben, aber keine Signifikanz aufwiesen.

Dieser Abschnitt beschränkt sich auf die Darstellung der Ergebnisse, die sich zur Bestätigung oder auch zur Widerlegung der aufgestellten Hypothesen eignen und Aussagen über die Auswirkungen des Einsatzes eines anthropomorphen Interface-Agenten in einem sprachbedienbaren Fahrerinformationssystem zulassen. Nach der Darlegung der Leistungsdaten der Benutzertests soll der folgende Abschnitt die Ergebnisse der subjektiven Bewertungen präsentieren, bevor in *Kapitel 5* anschließend diese Erkenntnisse den zu prüfenden Hypothesen gegenübergestellt werden.

#### 4.4.3.2. Post-Task-Fragebogen

Der Fragebogen, den die Probanden im Anschluss an die Arbeit mit den Prototypen beantwortet haben, gibt nach dem beobachteten Verhalten der Benutzer (*Task Performance*) den Blick auf deren eigene Einschätzung zu den von ihnen bedienten Systemen frei. Die Bewertungsskala erstreckte sich von der vollen Zustimmung zu negativ formulierten Aussagen über Neutralität zu der vollen Zustimmung zu entsprechenden, positiven Gegenüberstellungen. Für die Auswertung wurde die Siebener-Skala auf die Zahlenmenge  $B = \{-3; -2; -1; 0; 1; 2; 3\}$  abgebildet. So konnten arithmetische Funktionen wie z.B. die Bildung des Mittelwerts auf die einzelnen Items angewandt werden, deren Ergebnisse die Menge  $C = \{-3; \dots; 3\}$  ergab. Ist der Mittelwert aller Antworten eines Items z.B. 2,33 so bedeutet dies, dass die Probanden mit den positiven Aussagen sehr übereinstimmten und nur geringfügig von der vollen Zustimmung abwichen. Ein negativer Wert hingegen signalisiert je nach Betrag die Zustimmung zu einer negativen Aussage über das untersuchte System.

Im ersten Item wurden die Benutzer zur empfundenen Usability des Systems befragt. Aufgeschlüsselt auf die einzelnen Aussagen ergaben die Antworten aller Teilnehmer die folgenden Mittelwerte (Standardabweichung 0,87) (s. *Abb. 4.19*):

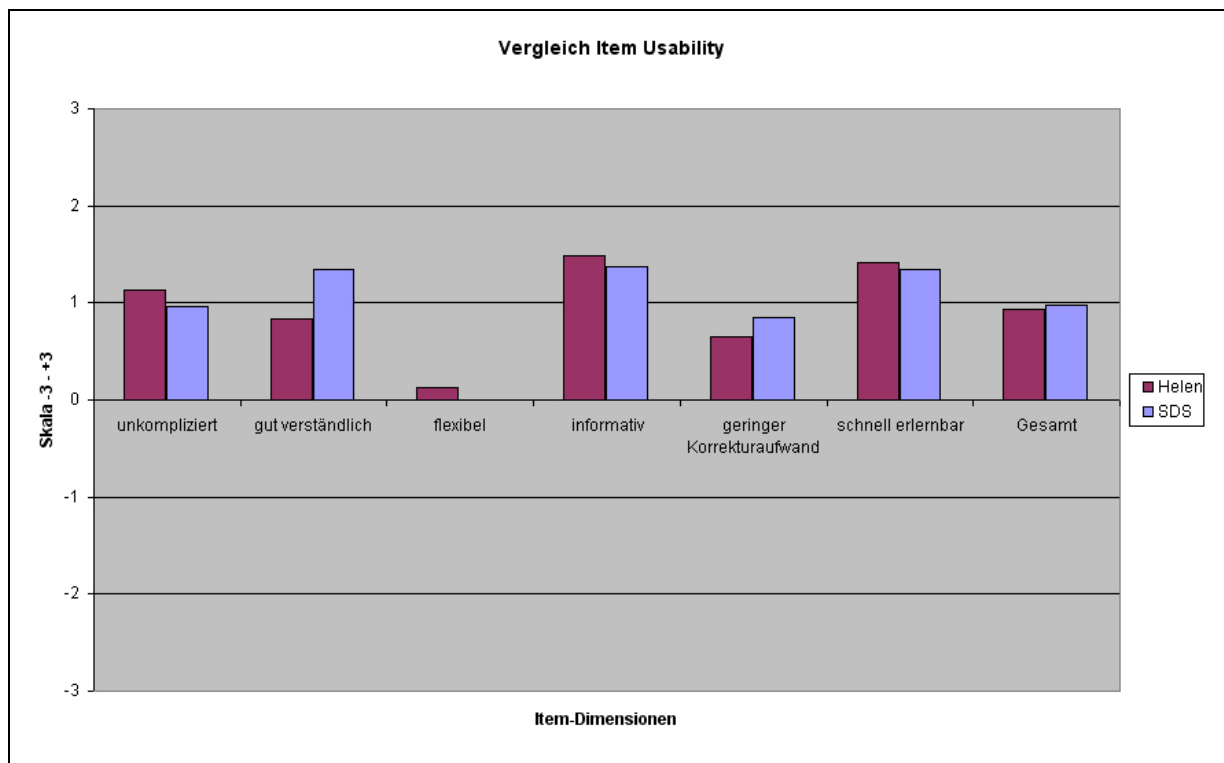


Abbildung 4.19: Fragebogen Item Usability

Beide Prototypen wurden also leicht positiv in ihrer Bedienbarkeit empfunden. In keiner der Usability-Dimensionen unterscheiden sich die Einschätzungen der Systeme grundlegend. Insgesamt ergab der Wert für dieses Item 0,94 für den Prototypen *Helen* und 0,98 für den *SDS Demonstrator*. Die mittlere Standardabweichung der einzelnen Aspekte lag jedoch zwischen 1,30 und 1,68. Das zeigt, dass die Probanden für ihre individuellen Bewertungen durchaus die gesamte Breite der Skala nutzten, sich aber insgesamt keine signifikanten Unterscheidungen der Systeme feststellen lassen.

Auch das Gesamtbild des nächsten Items erscheint recht einheitlich. (s. Abb. 4.20) Die Befragten zeigten bezüglich der subjektiven Empfindungen während der Benutzung des jeweiligen Systems keine grundsätzliche Abweichung in einer der Emotions-Dimensionen.

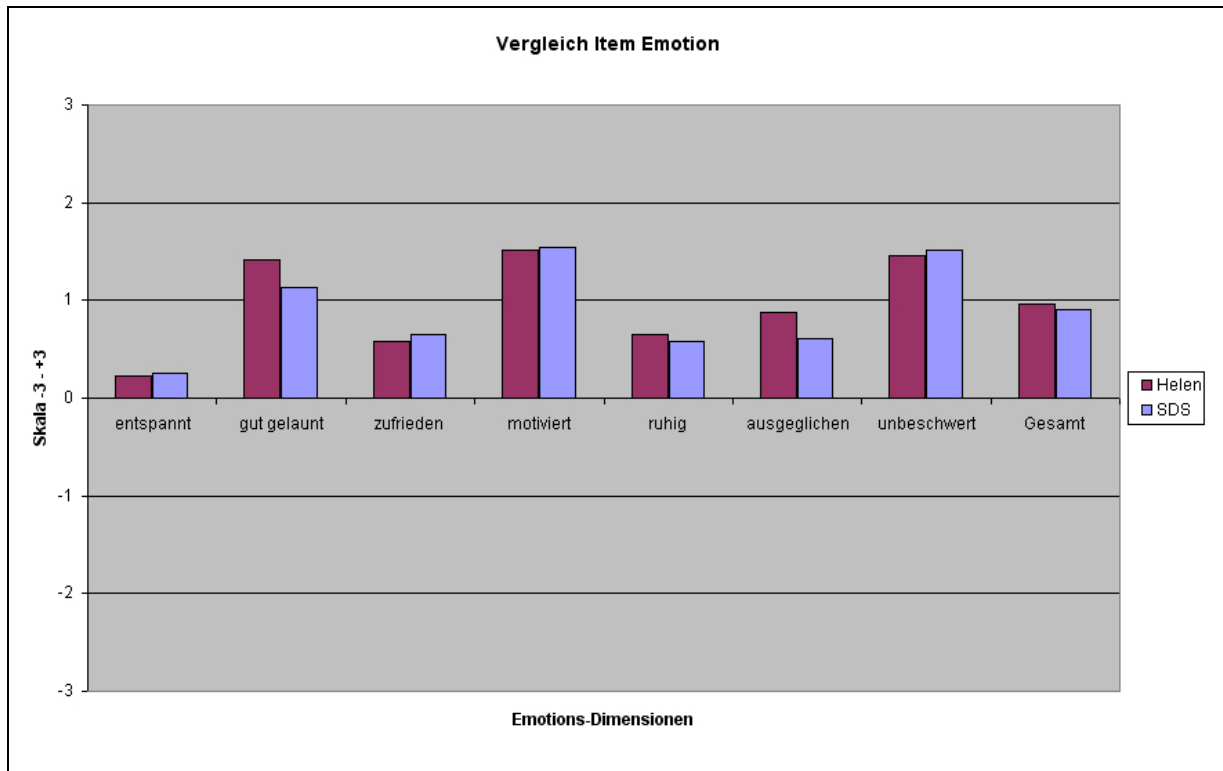


Abbildung 4.20: Fragebogen Item Emotion

Auch hier lag das Mittel der Antworten im leicht positiven Abschnitt der Siebener-Skala. Insgesamt ergab die Befindlichkeit während der Benutzung des Systems *Helen* einen durchschnittlichen Wert von 0,96 (Standardabweichung 0,87). Die Benutzer stimmten also geringfügig eher mit dem positiven Attribut der BASTI-Skala überein. Leicht schwächere positive Zustimmung (0,89) gaben die Probanden ab, die mit dem Prototypen *SDS* arbeiteten. Die größten Abweichungen der beiden untersuchten Gruppen traten bei den Attributen *gut gelaunt* und *ausgeglichen* auf. Hier schien die *Helen-Gruppe* mit jeweils ca. 0,2 positiver eingestellt als die Vergleichsgruppe der Probanden, die ohne einen virtuellen Charakter gearbeitet haben. Generell lässt sich aber feststellen, dass bei 31 Aussagen zur subjektiven Gemütslage während der Benutzung der Sprachdialogsysteme ähnliche Tendenzen bei beiden Prototypen zu erkennen sind.

Das dritte erfasste Item des Fragebogens ist eine abschließende Gesamtbewertung der Systeme. Nach Bewertung einzelner Aspekte hatten die Probanden die Gelegenheit, auf der ihnen bekannten Skala einzuordnen, wie ihnen der Prototyp insgesamt gefallen hat. *Abbildung 4.21* verdeutlicht zusammenfassend die geringen



Unterschiede in der Gesamtbewertung des Systems allgemein (Standardabweichung 1,04), der Usability (Standardabweichung 0,98) und der Stimmung der Probanden während der Benutzung (Standardabweichung 1,35).

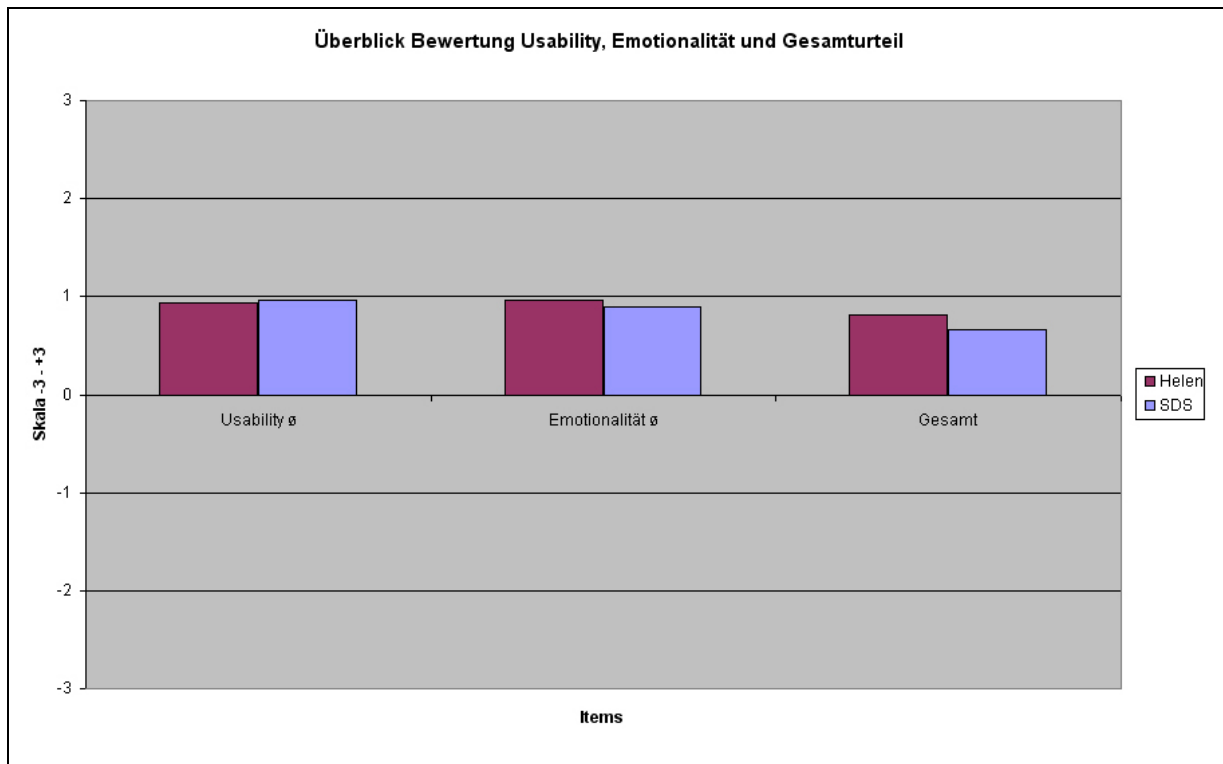


Abbildung 4.21: Überblick Gesamtbewertung Items 1 - 3

Eine geringfügige Tendenz dieses Gesamtüberblicks ist die leichte Besserstellung des *SDS Demonstrators* in der Bewertung der System-Usability (Differenz +0,33), wohingegen Probanden während der Benutzung des Prototypen *Helen* emotionale Attribute etwas positiver bewerten (Differenz: +0,06). Zusätzlich erhielt das animierte Sprachdialogsystem eine leicht positivere Gesamtbewertung (+ 0,04). Bei der untersuchten Stichprobe lassen sich aber insgesamt keine signifikanten Unterschiede in der Bewertung der Systeme bzw. der eigenen Stimmung während der Benutzung beobachten.

Diese Frage stellt das Ende der numerischen Bewertungen dar. Der folgende Teil stellte offene Fragen und ermöglichte den Probanden, ihre Meinung in freier, schriftlicher Form wiederzugeben. Diese Gelegenheit wurde intensiv genutzt und verschafft neben den Bewertungen auf der Skala einen unmittelbaren Eindruck

davon, was den Teilnehmern im Einzelnen an den Prototypen besonders gefiel bzw. was sie gestört hat. Zunächst fällt eine grundlegend positive Stimmung aller Probanden auf, den Einsatz der Sprachbedienung allgemein als vorteilhaft zu empfinden.

*„Das System war überraschend klar zu bedienen. Es bietet klare Vorteile vor allem bei der Texteingabe (Navigation: Strasse) gegenüber konventionellen Geräten.“ (SDS)*

*„Nach einer gewissen Eingewöhnungszeit ist es sicher ein guter Beitrag zu Verkehrssicherheit und Fahrkomfort.“ (Helen)*

Diese Bemerkungen lassen sich jedoch meist auf beide Prototypen gleichermaßen anwenden. Dadurch, dass einige Teilnehmer zum ersten Mal mit einem solchen System interagierten, bezogen sich ihre Kommentare oft auf das Konzept des Sprachdialogs im Auto an sich. Bei der Frage, was ihnen an dem Prototypen *Helen* und dem Dialog mit einem virtuellen Charakter besonders gefallen habe, antworteten zwei Probanden:

*„Je länger man sich damit beschäftigt und je besser man sich damit auskennt, desto mehr Spaß macht es.“*

*„Die aufmerksame Dame! Ich vermute aber, dass sie über kurz oder lang nervt.“*

Der Raum für Kritik an den Systemen wurde rege genutzt. Die Teilnehmer hatten fast alle innerhalb des Aufgabenszenarios eine oder mehrere kritische Situationen erlebt. Auf die Frage, was ihnen am sprachbedienbaren Fahrerinformationssystem nicht gefallen hat, antworteten beispielsweise zwei Teilnehmer wie folgt:

*„Manche Begrifflichkeiten waren unklar. Z.B. AutoStore für die Programmierung der Sender. Es war nicht immer klar über welche Tasten (oder Befehle) ich meine Aufgabe oder mein Ziel erreichen konnte.“*

*„Manchmal fehlte die Information über die nächsten Schritte oder Art der Fehler. Zudem waren lange Befehlseingaben nötig (kurze Befehle wären besser).“*

Gerade in Bezug auf System *Helen* war der Charakter selbst nicht nur Anlass für Lob und Gefallen, sondern auch Gegenstand der Kritik:

*„Die Assistentin ist etwas zu aktiv, präsent. Piepton muss bis zur nächsten Eingabe abgewartet werden.“*

*„Die Dame, die da rumhampelt, irritiert mich. Sie hilft nicht weiter, ist nicht schön anzusehen & bewegt ihre Lippen nicht zur Sprache.“*

Wenn dieser letzte Kommentar auch die stärkste negative Reaktion gegenüber *Helen* darstellt, wurde die Art und Weise der Darstellung doch oft kritisiert.<sup>19</sup> Diese beispielhaften Antworten im freien Teil des Items *Gesamturteil* der Systeme machen deutlich, dass die Systeme jeweils aus zwei unterschiedlichen Perspektiven beurteilt wurden. Eine ist die sachliche Ebene, die sich damit beschäftigt, inwiefern die Probanden mit der Spracherkennung und dem Konzept multimodaler Fahrerinformationssysteme zurechtgekommen sind. Hier gab es häufig Lob für den technologischen Ansatz, aber Kritik in Bezug auf die Terminologie und Dialogführung, welche gerade in den kritischen Situationen zu Fehlern führte. Eine weitere Perspektive betrachtet die visuelle Darstellung in Form von *Helen* oder der bloßen Befehlslisten. Die Kritik teilte sich in allgemeine Ablehnung und Details der Visualisierung bzw. des Einsatzes. Lob gab es speziell für den Charakter selten. Nur wenige Probanden bemerkten, dass sie die Animation unterhaltsam fänden. Häufig wurde auf den Anwendungskontext und die Gefahr der Ablenkung verwiesen.

Neben der allgemeinen Bewertung der Systeme sollten im darauf folgenden Abschnitt des Fragebogens die oben angesprochenen kritischen Situationen und das Motivations- bzw. Hilfe-Potenzial der Prototypen beleuchtet werden. Item 4 (*Difficulty*) fragte nach Schwierigkeiten in der Aufgabenbearbeitung und der empfundenen Hilfestellung des Systems. *Tabelle 4.3* zeigt zunächst, welche Aufgaben als besonders schwer zu erfüllen erachtet wurden.

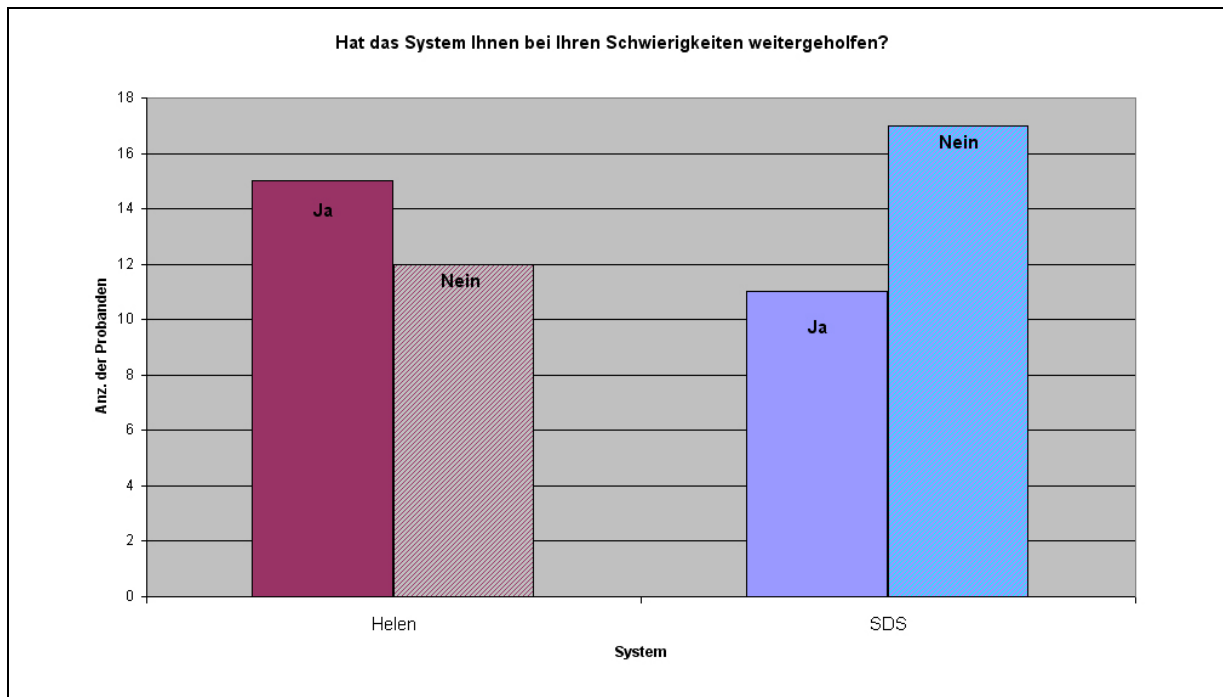
Welche Aufgaben waren besonders schwer zu erfüllen? (Mehrfachnennungen möglich)	
Telefonnummerneingabe	24
Radiosender automatisch speichern	11
Stadteingabe	10
Straßeneingabe	8
Falsche Eingaben im Navigationssystem korrigieren	8
Hausnummerneingabe	7
Falsche Eingabe im Telefon korrigieren	5
Radiosender wechseln und abspeichern	4
Maßstab der Kartenansicht ändern	3
CDs und Lieder wechseln und auswählen	3
Lieder automatisch anspielen lassen	2

**Tabelle 4.3: Item 4 – Teil 1 / Schwierigkeiten der Benutzer**

---

<sup>19</sup> Eine umfassende Darstellung der einzelnen Kommentare der Probanden befindet sich in *Anhang D*.

Die meistgenannten Aufgabenstellungen entsprechen der in der Konzeption des Benutzertests angestrebten Klimax kritischer Situationen im Mittelteil der Aufgabenszenarien. Die Benutzer zeigten die von den Experten erwarteten Schwierigkeiten in der Erledigung der genannten Aufgaben und bewerteten diese im Nachhinein entsprechend. Der entscheidende Aspekt der hier untersuchten Dimension ist die in solchen Situationen wahrgenommene Hilfestellung durch das System. Die anschließende Frage sollte deshalb untersuchen, inwiefern die Probanden eventuelle systemspezifische Unterschiede bei der Hilfestellung durch die Prototypen empfanden. *Abbildung 4.22* macht deutlich, dass tatsächlich die Hilfefunktion von System zu System anders wahrgenommen wurde.



**Abbildung 4.22: Item 4 - Teil 2 / Hilfefunktionen im Vergleich**

Probanden, die mit Prototyp *Helen* gearbeitet haben, antworteten zu 48,39% (15) positiv auf die Frage, ob Ihnen durch Systemhinweise in kritischen Situationen weitergeholfen wurde. 38,71% (12) hingegen bot das Dialogsystem keine Hinweise zur Lösung ihrer Bedienprobleme. Im Vergleich dazu fielen die Antworten der Teilnehmer, die den *SDS Demonstrator* bewerteten, negativer aus. Lediglich 35,48% (11) erhielten ihrer Meinung nach hilfreiche Unterstützung im Sprachdialog, 54,84%

(17) der Teilnehmer jedoch verneinten diese Frage. Während die Probanden bei der Interaktion mit einem virtuellen Charakter mehrheitlich eine Hilfestellung durch das System bemerken, fehlte ihnen ohne eine visualisierte Person im Referenzsystem eine solche Unterstützung. Vor diesem Hintergrund ist noch einmal zu betonen, dass beide Systeme sich in den Hilfefunktionen nur in der Hinsicht unterscheiden, wie der vorgelesene Text visualisiert wird, inhaltlich und funktional aber sind sie vollkommen äquivalent. Diese Unterschiede in der Wahrnehmung der Systemunterstützung in kritischen Situationen sind zwar in ihrer Tendenz deutlich erkennbar, aber mit einer Irrtumswahrscheinlichkeit von 0,226<sup>20</sup> nicht signifikant.

Die beiden abschließenden Kategorien des Fragebogens stellen das für den Benutzer aus den vorherigen Überlegungen resultierende Resümee dar. Unter Berücksichtigung der zu Papier gebrachten Erfahrungen und Bewertungen der beiden unterschiedlichen Prototypen sollten die Teilnehmer eine eventuelle Kaufabsicht äußern und urteilen, welches System ihnen mehr zusagt. Item 5 lässt keine weiteren Schlussfolgerungen zu, da jeweils diejenigen, die System 1 kaufen würden auch bereit wären in das Vergleichssystem zu investieren. Die Befragten scheinen also bei der Kaufentscheidung sich auf die generelle Funktion eines sprachbedienbaren Fahrerinformationssystems konzentriert und die unterschiedlichen Visualisierungen bei Preis und Kauf nicht bewertet zu haben. Jeweils 61% der Probanden (19) können sich vorstellen eines der beiden Produkte in ihrem Auto später einmal einzusetzen, 39% (12) verneinten diese Frage. Bei der abschließenden Beurteilung der Präferenz herrschte ein eindeutiges Urteil, das bei einer Irrtumswahrscheinlichkeit von 0,001 höchst signifikant ist. Auf die Frage, welches System sie generell für den Einsatz in ihrem Fahrzeug bevorzugen würden, antworteten 65% der Teilnehmer (20), dass sie ein Fahrerinformationssystem ohne virtuelle Informationsagentin bevorzugten. 23% (7) würden das Produkt mit AIA *Helen* wählen und jeweils 6% der Probanden (2) könnten sich beide gleichrangig oder gar nicht in ihrem Auto vorstellen. *Abbildung 4.23* zeigt die Aufteilung der Antworten unterteilt nach den Merkmalsunterscheidungen.

---

<sup>20</sup> Zur Berechnung der statistischen Irrtumswahrscheinlichkeit des Items 4 Teil 2 (*Abb. 4.22*) und des Items 5 (*Abb. 4.23*) wurde der Chi-Quadrat-Test nach Pearson angewandt.

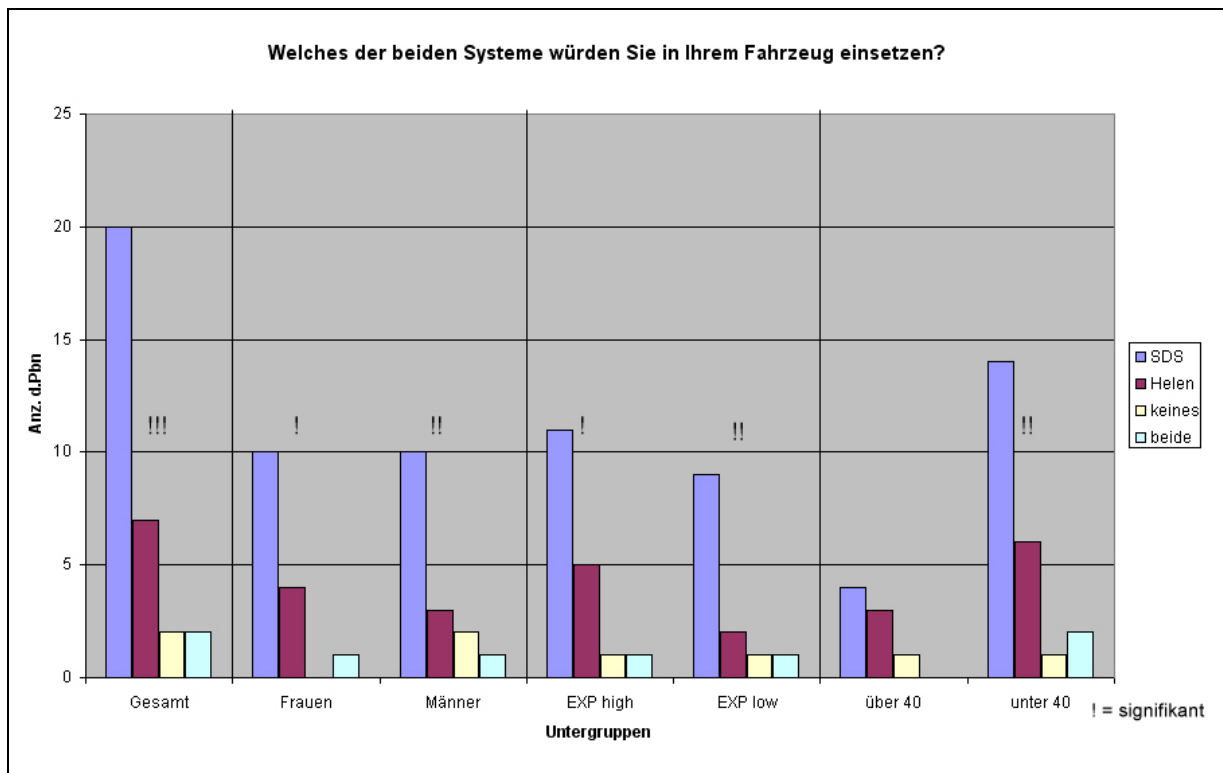


Abbildung 4.23: Präferenz der Systeme nach Merkmalsunterscheidungen

Es ist daher festzustellen, dass das Gesamtbild der Bevorzugung des reinen Sprachdialogsystems in seiner Tendenz in den Untergruppen der Stichprobe wieder zu erkennen ist. Jeweils 62,5% der Männer ( $p=0,01$ ) und 67% der Frauen ( $p=0,028$ ) gaben an, das Fahrerinformationssystem ohne Animation zu bevorzugen. Technisch erfahrene Anwender zeigten ebenfalls eine eindeutige Präferenz des *SDS Demonstrator*. Von ihnen sprachen sich 69% (11) für System *SDS* und 28% (5) für System *Helen* aus ( $p=0,04$ ). Teilnehmer mit geringen technischen Vorkenntnissen würden auch zu 69% (9) den *SDS Demonstrator* bevorzugt in ihrem Fahrzeug einsetzen ( $p=0,006$ ). Bei den jüngeren Probanden liegt die Ablehnung des Prototyps mit virtuellem Charakter bei 61% (14). Lediglich 22% (5) der Befragten würden System *Helen* in ihrem Fahrzeug einsetzen ( $p=0,018$ ). Die ältere Untersuchungsgruppe zeigt keine eindeutige Bevorzugung bzw. Ablehnung einer der beiden Prototypen (4 *SDS* / 3 *Helen*). Im Gegensatz zu den gemessenen positiven Leistungen mit System *Helen* zeigt sich demnach in der subjektiven Bewertung der Systeme eine signifikante Ablehnung des Prototypen mit anthropomorphem Interface-Agenten im Fahrerinformationssystem.

Die hier zusammengefasste Präsentation der quantitativ auswertbaren Ergebnisse stellt die Grundlage für die Schlussfolgerungen und Interpretationen des fünften Kapitels dar. Die ausführlichen Aussagen der Teilnehmer im freien Teil des Post-Task-Fragebogen befinden sich in *Anhang D* und ermöglichen qualitative Einblicke in die Erfahrungen der Benutzer.

#### 4.4.3.3. Physiologische Daten

Abschließend sollen zusätzlich die physiologischen Daten herangezogen werden, um sie den subjektiven Bewertungen gegenüberzustellen. Sie sollen über den messbaren Eindruck der emotionalen Verfassung der Probanden während der Interaktion mit den Systemen Aufschluss geben. Wie der Exkurs in *Kapitel 4.4.2.* bereits dargelegt hat, bestehen zur tatsächlichen Aussagekraft der erhobenen Werte unterschiedliche Ansichten. Vor allem sind die Methoden zur Messung von Emotionen durch Parameter wie Hautleitfähigkeit oder Herzfrequenz nicht eindeutig definiert. Es fehlt eine Klassifikation, die es erlaubt, bestimmten Messwerten einen Gemütszustand oder eine Empfindung zuzuordnen. Der oben dargestellte Ansatz wird hier exemplarisch untersucht. So werden die *EDA-Responses* und das *EDA-Level* grafisch ausgewertet und mit den Kennzahlen der *Task Performance* und den Bewertungen der Emotionalität der jeweiligen Probanden in Fragebogen-Item 2 gegenübergestellt. Dieser Abschnitt stellt also einen Ansatz zur Entwicklung von physiologischen Kennzahlen in Bezug auf Probanden-Emotionalität während der Benutzung dar. Es wird beispielhaft die Systematik dargestellt und eine abschließende Einschätzung zur Aussagekraft dieser Werte abgegeben.

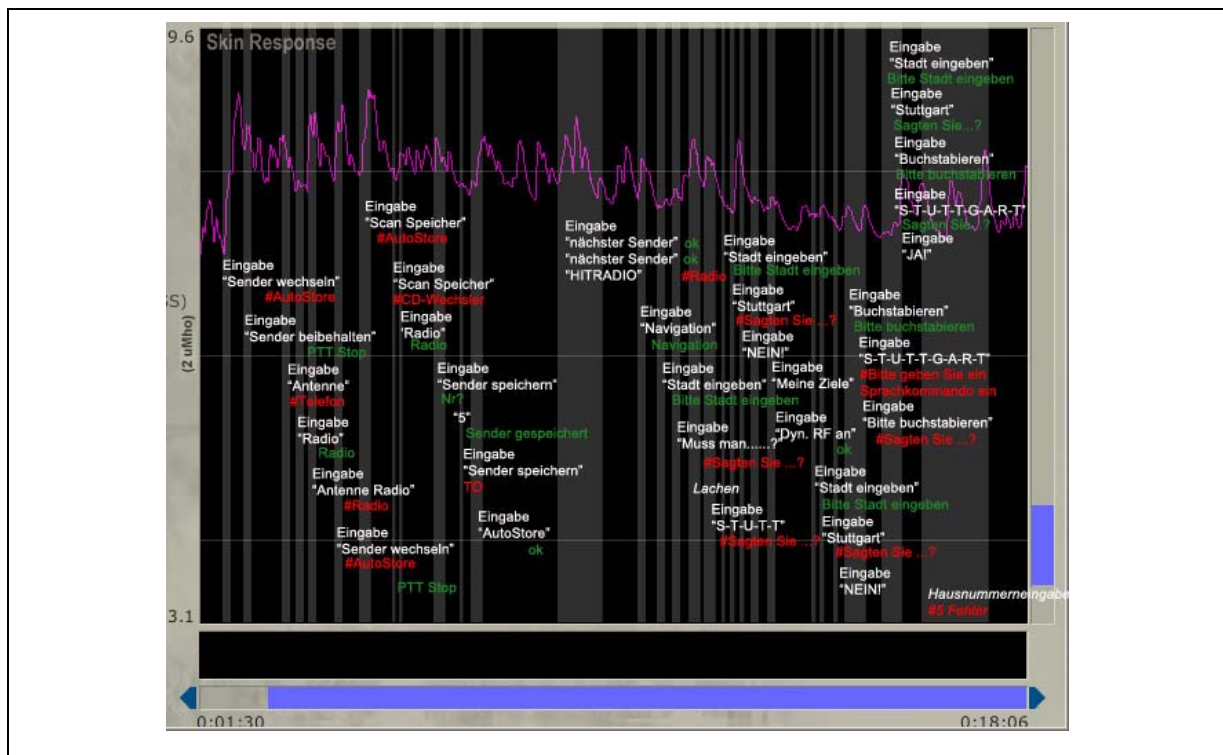
##### *Beispiel Probandin 042*

Probandin 042 ist Teil der Altersgruppe zwischen 20 und 30 Jahren und gab eine unterdurchschnittliche Erfahrung mit technischen Geräten an. Sie interagierte im ersten Teil des Benutzertests mit dem *SDS Demonstrator* und benötigte für die Bearbeitung der Aufgabenschritte 16:35 Minuten. In dieser Zeit unterliefen ihr 6 Eingabefehler, wobei sie fünf Mal ein nicht zur Verfügung stehendes Sprachkommando verwendete und einmal die Spracheingabe nicht rechtzeitig

ausführte. Insgesamt bearbeitete sie 86% der Aufgaben erfolgreich und benötigte dafür 38 Eingaben, die bei einer Spracherkennungsrate von 89% lagen. Sie bewertete im anschließenden Fragebogen die Usability des Systems mit durchschnittlich 1,17, ihre persönliche Stimmung während der Bedienung mit 1,00. Den Prototypen bewertete sie insgesamt mit 1, lobte die Übersichtlichkeit und schnelle Eingewöhnung, bemängelte aber die teilweise fehlerhafte Spracherkennung:

*„Die Spracherkennung ist mangelhaft. Man gerät ins Stocken, fühlt sich hilflos und etwas verunsichert. (in manchen Fällen)“*

Diesen detaillierten Einblick in den Ablauf der Probandin 042 dokumentiert *Abbildung 4.24* mit einem Ablaufprotokoll in Form des EDA-Graphen.



**Abbildung 4.24: EDA Aufzeichnung Proband 042 Task 1 SDS**

Die Balken markieren die Eingabeschritte der Benutzerin und geben an, um welche Befehle es sich handelte und ob sie erfolgreich vom System erkannt wurden. Der Kurvenverlauf der elektrodermalen Aktivität macht deutlich, dass der von Lin et al. beschriebene Graphenverlauf nicht zwingend mit einem Frustrationsereignis einhergehen muss, vielmehr auch in der Motorik des Probanden begründet sein kann. So ist bei jeder sprachlichen Äußerung im Verlauf des Versuchs ein gewisser



Kurvenanstieg zu beobachten. Lin et al. untersuchten in ihrer Studie die emotionale Reaktion der Teilnehmer auf Ereignisse in einem Computerspiel. Eine sprachliche Interaktion war somit kein Störfaktor. Dennoch scheinen sich die phasischen Reaktionen in ihrer Stärke zu unterscheiden. Um eine bessere Einschätzung der Bedeutung dieser Werte zu erhalten, werden im Folgenden *EDA-Responses* und – *Level* erfasst und exemplarisch an Proband 008 analysiert.

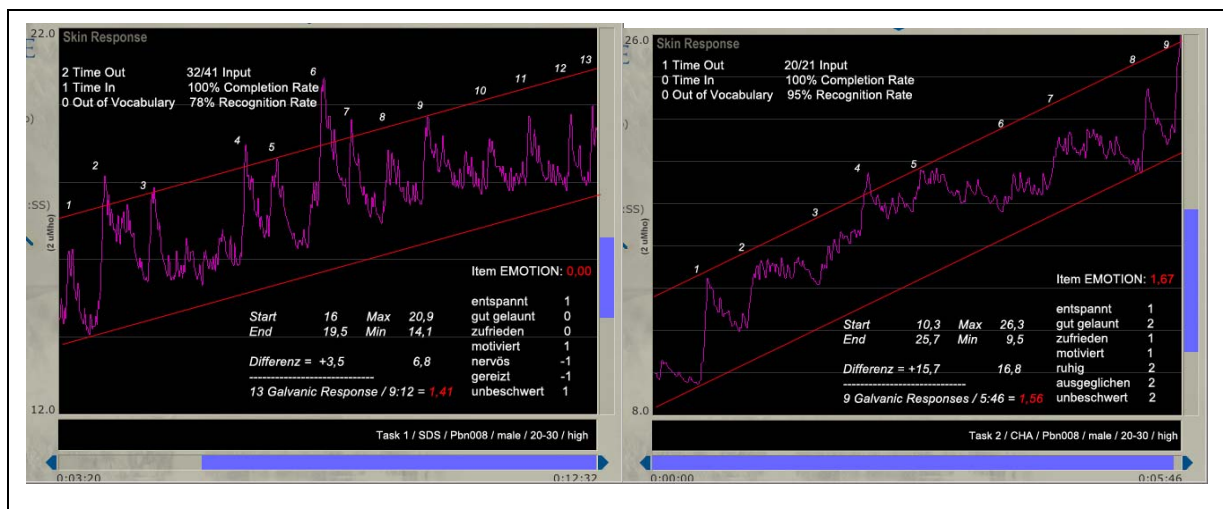


Abbildung 4.25: Physiologische Kennzahlen Pbn008 im Vergleich

#### Beispiel Proband 008

Die hier erfassten Daten sind also

- die Anzahl der phasischen Reaktionen pro Minute (*Resp./min*)
- die Differenz des Anfangs- und Endniveaus (*Diff. EDA Level*)
- die Differenz der Maximal- und Minimalwerte (*Diff. EDA Max Level*).

Sie werden mit den Erkenntnissen der Leistungsdaten und des Fragebogens verglichen:

- Bewertung des Item 2 Emotion des Fragebogens
- *Task Completion Rate (TCR)*
- *Recognition Rate*

<i>Pbn008</i>	System SDS	System Helen
Response / min	1,41	1,56
Diff. EDA Level	3,5	15,7
Diff. Max EDA Level	6,8	16,8
Item Emotion	0	1,67
TCR	100%	100%
Recognition Rate	78%	95%

Tabelle 4.4.: Vergleich der physiologischen Kennzahlen Pbn008

Während des ersten Aufgabenteils arbeitete Proband 008 mit dem SDS Demonstrator. Er löste alle Aufgaben erfolgreich in 9:12 Minuten bei einer Spracherkennungsrate von 78%. Er bewertete seine Stimmung während der Benutzung des ersten Systems als leicht nervös bzw. gereizt und erzielte damit einen Mittelwert des Items *Emotion* von 0,00. Die Anzahl der *EDA-Responses* lag bei 1,41 pro Minute, die Differenz des *EDA-Levels* betrug +3,5 und *Diff. Max Level* 6,8.

Betrachtet man diese Kennzahlen im Vergleich zu denen der Arbeitsphase mit System *Helen*, lässt sich eine erste Aussage treffen, inwiefern diese Werte miteinander korrelieren und eine Aussage bezüglich messbarer Emotionalität der Benutzer ermöglichen. Die zweite Bearbeitungsphase verlief deutlich positiver als die erste. Der Proband löste wieder 100% der Aufgabenstellung; diesmal bei einer Spracherkennungsrate durch Prototyp *Helen* von 95%. Diese fast reibungslose Bedienung des Systems schlägt sich unmittelbar in der Bewertung der Gemütslage des Benutzers wieder. Im anschließenden Fragebogen bewertete er das entsprechende Item mit 1,67 deutlich besser als nach dem ersten System. Die Anzahl der Reaktionen des EDA-Werts lag jedoch bei 1,56 pro Minute und das tonische *EDA-Level* zeigt einen konstanten Anstieg mit einer Differenz von 15,7. Damit liegen beide Werte deutlich über denen des ersten Systems. In diesem Fall scheint eine positivere Stimmung des Probanden eher positiv mit den aufgestellten Kennzahlen zu korrelieren. Die ursprüngliche Vermutung, dass ein phasischer Kurvenausschlag Anzeichen negativer Emotionen ist, trifft hier also nicht zu. Boucsein hingegen formulierte bereits sehr neutral, dass ein Anstieg des tonischen EDA-Werts lediglich ein Zeichen höherer emotionaler Beanspruchung sei. Eine Wertung, ob diese positiver oder negativer Art sei, konnten seine Ergebnisse nicht prüfen. Die Analyse

drei weiterer Probanden (s. *Anhang E*) ergab keine eindeutige Tendenz der untersuchten Korrelationen.<sup>21</sup> Die Vergleiche dieser Werte schwankten stark und wiesen keine eindeutigen Zusammenhänge auf. So standen teilweise gleichen Mittelwerten des Fragebogen-Items stark abweichende physiologische Messwerte gegenüber.

Es ist also festzuhalten, dass die aufgestellten Kennzahlen keine eindeutigen Erkenntnisse liefern bzw. die oben angesprochene Problematik der Sprachbedienung einen zu starken Störfaktor für die Erhebung physiologischer Werte darstellt. Insgesamt gesehen ist wohl die Kombination beider Aspekte entscheidend. Der Einfluss von sprachlichen Äußerungen auf den EDA-Wert ist eindeutig erkennbar. Daher kann die reine Betrachtung der Anzahl phasischer Reaktionen kein sinnvoller Messwert sein. <Beheben ließe sich diese Störung gegebenenfalls durch eine Ausgleichskurve, die den Graphenverlauf mittelt. Von diesem Mittelgraphen könnte der Grad der *EDA-Response* gemessen werden, der eine eventuelle Erkenntnis darüber liefert, ob die emotionale Reaktion von einem System zum anderen variiert bzw. mit den Aussagen des Fragebogens korreliert. Die erhobenen physiologischen Werte lassen einen Zusammenhang der elektrodermalen Aktivität mit der Emotionalität der Benutzer vermuten. Inwiefern aber diese Werte eine positive oder negative emotionale Reaktion beschreiben und welche Korrelationen zu den verwendeten Systemen bestehen, gilt es in weitergehenden Untersuchungen zu analysieren. Die exemplarisch aufgestellten Kennzahlen können dabei als ein erster Hinweis dafür gelten, wie Emotionalität von Probanden eventuell auch quantitative analysiert werden kann. Für die in *Kapitel 5* folgende Interpretation der Ergebnisse und Prüfung der Hypothesen können diese Werte jedoch nicht berücksichtigt werden.

---

<sup>21</sup> Es wurden die physiologischen Daten von drei Teilnehmern unterschiedlicher Probandengruppen grafisch analysiert. Eine allgemeine Tendenz möglicher Korrelationen oder Interpretationsmöglichkeiten war dabei nicht zu erkennen.

## 5. Auswertung und Interpretation der Ergebnisse

Die in *Kapitel 4* dargestellten Ergebnisse der heuristischen Evaluierung und des anschließenden vergleichenden Benutzertests zeigen gewisse signifikante Unterschiede in der Bewertung der beiden untersuchten Systeme aber auch bezüglich der Leistungsdaten der Probanden auf. Dieses Kapitel führt die einzelnen beobachteten Aspekte zusammen, um die eingangs formulierten Hypothesen zu prüfen, bevor dann Schlussfolgerungen für den Einsatz anthropomorpher Interface-Agenten in einem multimodalen Fahrerinformationssystem gezogen werden können.

### 5.1. Überprüfung der Hypothesen

Arbeitshypothese 1 basierte auf der wissenschaftlichen These<sup>22</sup>, dass bei der Gestaltung von Sprachdialogsystemen eine Anpassung der Mensch-Maschine-Interaktion an die zwischenmenschliche Kommunikation zu einer höheren Usability führt. So war die eingangs formulierte Vermutung, dass durch den Einsatz eines anthropomorphen Interface-Agenten im Sprachdialog eines Fahrerinformationssystems die Leistungsdaten verbessert werden können. Die Ergebnisse des Benutzertests stützen diese Hypothese. Probanden, die mit dem System *Helen* das erste Aufgabenszenario bearbeiteten, benötigten signifikant weniger Zeit und waren zu einem höheren Grad erfolgreich als Teilnehmer, die ohne einen anthropomorphen Interface Agenten arbeiteten. Bei genauerer Betrachtung der Daten wird dabei deutlich, dass dieser Unterschied bei den männlichen Probanden nicht so ausgeprägt ist wie bei den weiblichen. Durchgängig ist aber besonders die bessere Leistung der Probanden mit wenig technischer Erfahrung zu bemerken, wenn sie mit einem virtuellen Charakter interagierten. Nach Eintreten des Lerneffekts sind diese Unterschiede aber nur noch minimal vorhanden. Die subjektiven Bewertungen der Probanden lassen für diese Hypothese keine Schlussfolgerungen zu. Es ist lediglich zu bemerken, dass die Ergebnisse der Fragebögen keine direkten

---

<sup>22</sup> Siehe hierzu CASSELL 2001, KRÄMER / BENTE 2002, RIST et al. 2004

Unterschiede in der Usability bzw. allgemeinen Bewertung der Systeme erkennen ließen.

Arbeitshypothese 2 war Schlussfolgerung der Beobachtungen anderer Einsatzbereiche virtueller Charaktere. Die Erfahrung hatte gezeigt, dass gerade als Marketing-Instrument anthropomorphe Verkäufer einen hohen Zuspruch erhielten<sup>23</sup>. Diese Beobachtung konnte im Einsatz der vorliegenden multimodalen Fahrerinformationssysteme nicht gemacht werden. Die erwartete erhöhte Ausgabebereitschaft für das System mit einem virtuellen Charakter ist nicht zu beobachten. Im Gegenteil, die Probanden lehnten das System *Helen* mit 62-prozentiger Mehrheit ab. Lediglich 26% der Befragten können sich vorstellen, in ihrem Fahrzeug ein sprachbedienbares Fahrerinformationssystem mit einer virtuellen Informationsagentin einzusetzen. Diese Beobachtung verteilt sich gleichmäßig auf die einzelnen Untergruppen der Stichprobe. So stehen also die Ergebnisse der Benutzerleistungen in starkem Gegensatz zur subjektiven Akzeptanz der Probanden.

Arbeitshypothese 3 beschäftigte sich mit der Usability-Richtlinie der Nutzermotivation und –unterstützung. Eine Benutzerschnittstelle soll laut Norm (DIN 14915 Teil 1) den Nutzer motivieren können. Dies konnte gerade in dieser Untersuchung durch die Erzeugung kritischer Situationen mit hohem Fehlerpotenzial geprüft werden. *AH 3* erwartete also, dass in solchen Situationen eine niedrigere Frustration der Nutzer durch bewusster empfundene Hilfestellung seitens des Systems *Helen* eintreten würde. Ein niedrigeres Frustrationsniveau ist aus den Ergebnissen des Post-Task-Fragebogens nicht zu erkennen. Die Teilnehmer bewerteten ihre Stimmung während der Benutzung der Systeme gleichwertig. Die erhoffte Prüfung dieser subjektiven Daten durch die quantitative Messung physiologischer Kennzahlen erwies sich leider nicht als aussagekräftig. In der abschließenden Benutzerbefragung urteilten die Probanden jedoch eindeutig: Bei der Interaktion mit einem anthropomorphen Interface-Agenten empfanden sie zu einem größeren Anteil Hilfeleistungen als bei dem Referenzsystem. Diese Erkenntnis stützt also Arbeitshypothese 3 und weist

---

<sup>23</sup> siehe hierzu STRICKER 2003

daraufhin, wie eine anthropomorphe Schnittstelle gerade in Fehlersituationen unterstützend wirken kann.

Die abschließende Arbeitshypothese 4 äußerte als Resultat der heuristischen Evaluierung die Vermutung, dass ein AIA das Potenzial aufwiese, besonders in den Kategorien *Selbstbeschreibungsfähigkeit* und *Motivation* einen Mehrwert zur Usability eines sprachbedienbaren Fahrerinformationssystems beizutragen. Die tatsächlich gemessene Leistungssteigerung und empfundene Hilfeleistung des Systems *Helen* ist bereits in den Thesen *AH1* und *AH3* bestätigt. Jedoch äußerten die Experten den Hinweis, dass dieses Potenzial im Fahrzeug nur ausgenutzt werden könnte, wenn „die Gestaltung der Interaktion mit dem virtuellen Charakter auf Sicherheitsrichtlinien im Fahrzeug abgestimmt ist und so eine empfundene Ablenkung von der Primäraufgabe des Fahrers vermeidet.“ (s. Seite 49) Inwiefern eine Schnittstelle Sicherheitsrichtlinien für die Interaktion im Fahrzeug erfüllt, wurde in dieser Untersuchung nicht eigens geprüft. Die Aussagen der Probanden geben aber einen direkten Hinweis:

*„Ich benötige die Frau zur Ansprache nicht. Sie lenkt eher vom Verkehr ab.“*

*„Ich fand den Avatar zu ablenkend.“*

*„Beide sind von der Bedienung her gleich. Beim 2. System irritiert mich die Person auf dem Display - man wird mehr abgelenkt, man konzentriert sich auf die Bewegungen der Frau.“*

Ein großer Anteil der Probanden begründete die Ablehnung des Prototypen *Helen* mit dem Argument, dass die Animation irritierend sei und den Blick von der Straße ablenke. Diese durchgängig beobachteten Aussagen stützen nicht nur direkt Arbeitshypothese 4, sondern bieten somit auch eine Erklärung für die überraschende Widerlegung der zweiten Hypothese und der geringen Benutzerakzeptanz des anthropomorphen Interface-Agenten.

Die Prüfung der Hypothesen ergibt also ein differenziertes Bild. Die einzelnen Erkenntnisse der Arbeitshypothesen gilt es in einen Gesamtkontext zu stellen. Was lässt sich insgesamt aus der heuristischen Evaluierung und aus dem Benutzertest

schließen? Ist der Einsatz eines anthropomorphen Interface-Agenten in einem multimodalen Fahrerinformationssystem zu empfehlen? Und was gilt es bei der Implementierung eines solchen Dialogsystems zu berücksichtigen? Der folgende Abschnitt führt die Erkenntnisse dieser Arbeit zusammen und leitet Schlussfolgerungen für den Einsatz anthropomorpher Interface-Agenten im Fahrzeug ab.

### **5.2. Ursachenanalyse der beobachteten Ergebnisse**

Die eingangs aufgestellten Hypothesen richteten sich nach der ursprünglichen Motivation und Fragestellung dieser Arbeit: Bietet ein anthropomorpher Interface-Agent einen Mehrwert bezüglich Benutzerakzeptanz und Usability eines multimodalen Fahrerinformationssystems? Die Prüfung der Arbeitshypothesen beantwortet diese Frage nicht eindeutig. Zum einen sprechen die erhobenen Leistungsdaten der Teilnehmer des Benutzertests eine klare Sprache. Interagierten sie mit einer virtuellen Informationsagentin statt mit einer rein schematischen Darstellung des Dialogsystems, lösten sie die ihnen gestellten Aufgaben schneller, mit höherer Erfolgsquote und empfanden sie das System in kritischen Situationen mehrheitlich als hilfreich bei der Fehlerkorrektur. Ein direkt messbarer Mehrwert des Einsatzes eines anthropomorphen Interface-Agenten bezüglich der Usability in einem multimodalen Fahrerinformationssystem ist zunächst feststellbar.

Die subjektive System-Bewertung der Teilnehmer macht jedoch deutlich, dass Benutzerakzeptanz nicht automatisch Ergebnis der gemessenen Leistungsdaten ist. Die Probanden äußerten mehrheitlich, dass sie für die Interaktion mit einem Fahrerinformationssystem eher ein schematisches Interface bevorzugten. So wurde die Visualisierung in Form einer Informationsagentin als störend und ablenkend empfunden. Ein virtueller Charakter ist demnach kein entscheidender Kaufgrund. Ein Mehrwert bezüglich der Benutzerakzeptanz ist nicht zu beobachten.

Erst die Analyse der Ursachen dieser Ergebnisse ermöglicht Schlussfolgerungen darüber, ob ein anthropomorpher Interface-Agent tatsächlich in einem Fahrerinformationssystem eingesetzt werden kann und wie eine Implementierung

aussehen sollte, um einen Mehrwert bezüglich beider Aspekte, also der Usability und der Benutzerakzeptanz, zu erreichen. Warum also sind die Benutzer zwar messbar besser mit dem Fahrerinformationssystem zurecht gekommen, das durch einen virtuellen Charakter repräsentiert wurde, würden sich aber bei einer Kaufentscheidung gleichwohl mehrheitlich für das andere System ohne grafische Animation entscheiden?

Eine mögliche Begründung für die besseren Leistungsdaten in Zusammenhang mit der anthropomorphen Schnittstelle liefert die Beobachtung der Probanden während des Benutzertests. So war bei der Interaktion mit dem virtuellen Charakter *Helen* zu beobachten, dass die Probanden dem Display größere Aufmerksamkeit schenkten. Die Auseinandersetzung mit der Darstellung des Prototypen war aufgrund der visuell anspruchsvolleren Benutzerschnittstelle intensiver. Es lässt sich vermuten, dass die genauere Betrachtung des Displays dazu führte, dass die Benutzer die grafisch dargestellten, essentiellen Informationen - wie z.B. die Liste der zur Verfügung stehenden Sprachkommandos - mehr beachteten. Mit gesteigerter Aufmerksamkeit sank so die Gefahr, dass einige Befehle nicht wahrgenommen werden. Zusätzlich zu dieser Beobachtung der Benutzertests lassen sich Erkenntnisse anderer Studien mit anführen. So konnten auch Walker, Sproull und Subramani (1994) positive Leistungsveränderungen bei Nutzern feststellen, die mit einem synthetischen Gesicht arbeiten. Personen, die mit einem anthropomorphen Interface interagierten, ließen sich mehr Zeit, machten weniger Fehler und antworteten ausführlicher auf offene Fragen. (vgl. KRÄMER / BENTE 2002: 214) In der oben bereits erwähnten Studie *REA* des *MIT Media Lab* erläutert Cassell zudem, worin für sie die Ursachen solcher positiven Einflüsse auf die Mensch-Maschine-Interaktion bestehen: „Gaze, gesture, intonation, and body posture play an essential role in the proper execution of many conversational behaviors—such as conversation initiation and termination, turn taking and interruption handling, and feedback and error correction—and these kinds of behavior enable the exchange of multiple levels of representation of information in real time.” (2001:76) Diesen Studien und Meinungen zufolge sind also die positiven Leistungsveränderungen, die auch in dieser Arbeit beobachtet werden konnten, Resultat der effektiven Werkzeuge zwischenmenschlicher Kommunikation (Gestik,



Mimik, ...), die durch den Einsatz eines anthropomorphen Interface-Agenten die Mensch-Maschine-Interaktion positiv beeinflussen.

Worin liegt dann aber die Ursache für die geringe Benutzerakzeptanz des Systems mit der virtuellen Informationsagentin *Helen*? Und welche Schlussfolgerungen lassen sich aus diesen subjektiven Bewertungen ziehen? Die Ablehnung der Probanden hat zwei unterschiedliche Beweggründe. Zum einen äußerte sich die Mehrheit der Teilnehmer kritisch zu der konkreten Umsetzung des virtuellen Charakters. Viele konnten keine direkte Bedeutung in den ausgeführten Bewegungen erkennen. Einige bemängelten auch, dass die Lippen nicht synchron zum gesprochenen Text seien. Wiederum andere meinten, dass *Helen* teilweise wichtige Informationen im Hintergrund verdecken würde. Diese Kommentare der Benutzer zeigen, dass gerade die oben erwähnten sinntragenden Elemente nonverbaler Kommunikation in dem vorliegenden Prototypen keinem gesamtheitlichen Konzept zu folgen scheinen und einige Details in der Visualisierung des virtuellen Charakters noch nicht die volle Effektivität dieser Bewegungen nutzen. Tatsächlich fand eine kooperative Konzeption des Charamel-Charakters und seiner Funktion im Sprachdialogsystem der Blaupunkt GmbH nicht in dem Ausmaß statt, dass dem vorliegenden Prototypen ein umfassendes Bedienkonzept zugrunde läge. Als zweite Ursache der Ablehnung war vielfach die Meinung der Probanden zu beobachten, dass sie solch eine „Spielerei“ nicht benötigten; dass die aufwändige Animation im Fahrzeug keinen essentiellen Zweck erfülle. Die hohe Sensibilisierung der Fahrer für Sicherheit und Funktionalität, führt zu einer genauen Prüfung, ob denn wirklich eine solche Animation, die stets im Blickfeld präsent ist, unbedingt notwendig sei. Gerade diese Sicherheitsbedenken sind ein entscheidendes Kriterium für eine spätere Kaufentscheidung eines Fahrerinformationssystems. Der untersuchte Prototyp erfüllte der Bewertung der Teilnehmer zufolge diese speziellen Anforderungen des automobilen Umfelds nicht.

### **5.3. Schlussfolgerungen für den Einsatz anthropomorpher Interface-Agenten**

Aus dieser Analyse der Ursachen der beobachteten Reaktionen auf die untersuchten Prototypen lassen sich Schlussfolgerungen für den Einsatz anthropomorpher Interface-Agenten in einem multimodalen Fahrerinformationssystem ziehen.

Primäres Ziel der Gestaltung einer Benutzerschnittstelle im Fahrzeug ist es, dem Fahrer die Funktionen und Informationen so effizient und effektiv wie möglich zugänglich zu machen. Die Basis hierfür stellt zunächst eine Dialogschnittstelle dar, die konform zu bestehenden Usability-Guidelines entwickelt wurde. So geben gerade die Erkenntnisse der heuristischen Evaluierung (s. *Anhang B*) wichtige Hinweise, was in beiden Systemen bezüglich Darstellung und Dialogführung noch verbesserungswürdig ist. Dazu gehören vor allem die folgenden Grundsätze:

- *Terminologie*

Die Bezeichnungen der Befehle und Funktionen sollten sich stärker am alltäglichen Sprachgebrauch anpassen

- *Konsistenz*

Die verwendeten Befehle und Begriffe der Dialogführung sollten einheitlich und konsistent in allen Anwendungen verwendet werden

- *Systemstatus*

Dem Benutzer sollte ersichtlich sein, was das System eventuell falsch verstanden hat und was die Ursache dafür sein könnte

- *Konstruktive Hilfe*

Mit konstruktiver Hilfestellung sollte das System dem Benutzer dabei behilflich sein, wie er ein aktuelles Problem lösen kann

- *Individualisierung*

Dem Benutzer sollte soweit wie möglich, die Gelegenheit geboten werden, die Dialogsequenzen abzukürzen (Shortcuts) und die Bedienung an seine Wünsche anzupassen

Ein weiteres Ergebnis dieser Arbeit ist, dass der Einsatz eines anthropomorphen Interface-Agenten die Leistung der weiblichen und technisch unerfahrenen Benutzer wesentlich steigern kann. Zudem wird die Darstellung des Sprachdialogs durch einen virtuellen Charakter als hilfreich empfunden. Um dieses Potenzial nutzen zu können, gilt es jedoch die Benutzerakzeptanz zu steigern und bei der Optimierung des anthropomorphen Prototypen einige Richtlinien zu berücksichtigen:

- *Authentizität*

Den nonverbalen Elementen des Interface-Agenten sollte ein Konzept hinterlegt sein, das die Bedeutung jeder Bewegung definiert und bewusst einsetzt, um eine größtmögliche Authentizität des Charakters zu erzeugen.

- *Synchronität*

Die Lippenbewegungen und die Gestik des virtuellen Charakters sollten sich noch näher an der Sprachausgabe orientieren.

- *Funktionalität*

Die Potenziale des virtuellen Charakters gerade bei Hilfe-Funktionen durch Körpersprache Inhalte verständlicher zu vermitteln sollten weiter ausgebaut werden.

- *Benutzerkontrolle*

Dem Benutzer sollte jederzeit die Möglichkeit gegeben werden, die Animation auszublenden (evtl. automatisch während der Fahrt).

Verwendet man ein stimmiges Konzept der Darstellungsform und der Funktionalitäten eines anthropomorphen Interface-Agenten, kann man dessen Wirksamkeit steigern und die positiven Ergebnisse der Leistungsdaten auch auf die Nutzerakzeptanz übertragen. Gerade die Authentizität eines virtuellen Charakters ist für die Benutzerakzeptanz von großer Bedeutung. Die verfügbaren nonverbalen Mittel sind bedeutungsvolle Elemente der zwischenmenschlichen Kommunikation. Setzt man sie unbewusst oder gar falsch ein, bewirken sie eine starke Entfremdung der Benutzer gegenüber dem Charakter. Cassell et al. betonen ebenfalls die Bedeutung eines schlüssigen Konzepts, das jeder kommunikativen Funktion eine gezielte Bewegung zuordnet: „We have argued that embodied conversational agents are a logical and

needed extension to the conversational metaphor of human-computer interaction. We argue, however, that embodiment needs to be based on an understanding of conversational function, rather than a additive – and ad hoc – model of the relationship between nonverbal modalities and verbal conversational behaviors.“ (1999: 527).

Erstellt man also eine Art Bewegungskatalog und hinterlegt jeder Bewegung des virtuellen Charakters eine kommunikative Funktion, nutzt man das volle Potenzial anthropomorpher Schnittstellen erst aus und erhöht nicht nur die Authentizität des Interface-Agenten, sondern auch die Akzeptanz der Benutzer. Ein solches Bedienkonzept kann nicht aus der bloßen Kombination eines Sprachdialogssystems und einer zusätzlichen Animation bestehen, sondern bedarf einer gemeinsamen Spezifikation der Blaupunkt GmbH und der Charamel GmbH, die die Besonderheiten einer anthropomorphen Schnittstelle mit in die Dialoggestaltung einbezieht.

## 6. Fazit und Ausblick

Die Ausgangsfragen dieser Arbeit bezogen sich auf die Potenziale einer anthropomorphen Benutzerschnittstelle im Sprachdialog der Blaupunkt GmbH. Welcher Mehrwert kommt einem Fahrerinformationssystem mit menschlichem Antlitz zu? Vermag diese Technologie die Akzeptanz und Usability von Sprachdialogsystemen zu steigern? Wie reagieren die Benutzer auf einen virtuellen Charakter als Ansprechpartner ihres Informationssystems im Fahrzeug?

Um diesen Fragen auf den Grund zu gehen, wurden die vorliegenden Prototypen der Blaupunkt GmbH und der Charamel GmbH in drei Schritten analysiert. Zunächst diente die informationswissenschaftliche Recherche einem ersten Überblick, was in Bezug auf die Usability von Sprachapplikationen zu berücksichtigen ist und was für bisherige Erkenntnisse zum Einsatz von anthropomorphen Interface-Agenten vorliegen. Eine heuristische Evaluierung in Form von Experten-Interviews trug daraufhin dazu bei, kritische Elemente in der Bedienbarkeit des prototypischen Sprachdialogsystems der Blaupunkt GmbH zu identifizieren und potenzielle Auswirkungen des virtuellen Charakters *Helen* der Charamel GmbH einzuschätzen. Der aus diesen Erkenntnissen konzipierte Benutzertest ermöglichte schließlich in quantitativen und qualitativen Analysen die Benutzerakzeptanz und Usability einer schematischen und einer anthropomorphen Benutzerschnittstelle im Vergleich zu evaluieren.

Als Ergebnis dieser Arbeit lassen sich zwei Feststellungen konfrontieren. Zum einen ließ sich beobachten, dass die Probanden – im Besonderen Frauen und technisch Unerfahrene - mit einem anthropomorphen Interface-Agenten signifikante Leistungssteigerungen aufwiesen. Zum anderen zeigte die übergreifende Ablehnung des Fahrerinformationssystems mit virtuellem Charakter, dass sich bessere Leistungen nicht automatisch in gesteigerter Benutzerakzeptanz niederschlugen.

Beide Erkenntnisse lassen sich anhand eines Zitats Cassells analysieren:

„As I've argued, used *appropriately*, an embodied interface can provide not only something pretty and entertaining to look at but can also enable the use of certain communication protocols in face-to-face conversation that facilitate user interaction and provide for a more rich and robust channel of communication than is afforded by any other mediated channel available today.” (2001: 76)

Die bessere Usability der anthropomorphen Benutzerschnittstelle im Vergleich zum reinen Sprachdialog ist demnach durch die Simulation der zwischenmenschlichen Kommunikation begründbar. Der Benutzer betrachtet aufgrund der grafisch aufwändigeren Darstellung das Display aufmerksamer und bekommt dank der eingesetzten nonverbalen Mittel effektiver und effizienter die nötige Information vermittelt.

Einen entscheidenden Punkt spricht Cassell im obigen Zitat aber ebenso an: Die *Angemessenheit* der gewählten Gestaltung. Diese lässt sich auf die Wahl der Körpersprache des Charakters ebenso wie auf den Anwendungskontext des Autofahrens beziehen. Ein anthropomorphes Sprachdialogsystem sollte demnach einem schlüssigen Konzept für den Einsatz von Gestik und Mimik des virtuellen Charakters folgen und diesen nur so in die Interaktion mit dem Benutzer einbinden, dass die empfundene Sicherheit des Fahrers unter allen Umständen gewährleistet ist. Orientiert man sich in der weiteren Umsetzung eines multimodalen Fahrerinformationssystems an den daraus entstehenden Richtlinien, versprechen sich die in dieser Arbeit beobachteten Potenziale eines anthropomorphen Interface-Agenten nicht nur in der Usability sondern auch in der Benutzerakzeptanz niederzuschlagen.

# Literaturverzeichnis

## Bücher, Monographien und Nachschlagewerke

BALZERT, H.; HOPPE, H.U.; OPPERMAN, R.; PESCHKE, H.; ROHR, G.; STREITZ, N.A. (Hrsg.) (1988): *Einführung in die Software-Ergonomie*. Berlin; New York: Walter de Gruyter.

BOUCSEIN, W. (1988): *Elektrodermale Aktivität. Grundlagen, Methoden und Anwendungen*. Berlin et al.: Springer-Verlag.

DEUTSCHES INSTITUT FÜR NORMUNG e.V. (2004): *DIN Taschenbuch 354. Software-Ergonomie. Empfehlungen für die Programmierung und Auswahl von Software*. Berlin: Beuth Verlag GmbH. (CD-ROM)

DUDEN (2001): *Duden Band 5. Fremdwörterbuch*. Mannheim et al.: Dudenverlag.

DUMAS, J.S.; REDISH, J.C. (Hrsg.) (1999): *A practical guide to usability testing*. Norwood, NJ: Ablex Publishing.

DÖRING, N. (1999): *Sozialpsychologie des Internet. Die Bedeutung des Internet für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen*. Göttingen: Hogrefe Verlag für Psychologie

GIBBON, D.; MERTINS, I.; MOORE, R. K. (2000): *Handbook of multimodal and spoken dialogue systems: resources, terminology, and product evaluation*. Boston et al.: Kluwer Academic Publishers.

KRAUSE, J.; HITZENBERGER, L. (Hrsg.) (1992): *Computer Talk*. Hildesheim et al.: Georg Olms Verlag.

PEISSNER, M.; BIESTERFELDT, J.; HEIDMANN, F. (2004): *Akzeptanz und Usability von Sprachapplikationen in Deutschland*. Stuttgart: Fraunhofer IRB Verlag.

PREIM, B. (1999): *Entwicklung interaktiver Systeme: Grundlagen, Fallbeispiele und innovative Anwendungsfelder*. Berlin et al.: Springer-Verlag.

SHNEIDERMAN, B. (1998): *Designing the user interface: strategies for effective human-computer-interaction*. Reading, MA et al.: Addison-Wesley.

### **Artikel in Zeitschriften oder Sammelbänden**

AKYOL, S.; LIBUDA, L.; KRAISS, K.-F. (2001): *Multimodale Benutzung interaktiver Kfz-Bordsysteme*. In: JÜRGENSOHN, Th.; TIMPE, K.-P. (Hrsg.) (2001): *Kraftfahrzeugführung*. Berlin: Springer-Verlag, pp. 137-154.

CASSEL, J.; THÓRISSON, K. R. (1999). *The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents*. In: *Journal of Applied Artificial Intelligence*, 13 (3). Taylor & Francis, pp. 519-538. Verfügbar unter: <<http://taylorandfrancis.metapress.com/link.asp?id=bpnlwxfxbvmb1bt8>> (Zugriff am 21.03.2006, 14:36 Uhr)

CASSELL, J. (2001): *Embodied Conversational Agents. Representation and Intelligence in User Interfaces*. In: *AI Magazine. Winter 2001. Volume 22. Issue 4*. Menlo Park, CA: AAAI, pp. 67-84.

CHEYER, A.; JULIA, L. (1999): *InfoWiz: An Animated Voice Interactive Information System*. In: *Agents '99: workshop 6: Communicative Agents*. Verfügbar unter: <<http://bravobrava.com/people/julia/articles/agents99.pdf>> (Zugriff am 5.3.2006, 17:43 Uhr)



DYBKJAER, L.; BERNSEN, N.O. (2000): *Usability issues in spoken dialogue systems*. In: *Natural Language Engineering* 6 (3-4). Cambridge: Cambridge University Press, pp. 243-271.

HASSENZAHL, M.; HOFVENSCHIÖLD, E. (2003): *If it doesn't feel right, who cares if it works? Oder muss Software mehr als nur gebrauchstauglich sein?*. In: PEISSNER, M.; RÖSE, K. (Hrsg.) (2003): *Usability Professionals 2003*. Stuttgart: GC-UPA e.V., pp. 135-138.

HASSENZAHL, M.; SEEWALD, F. (2004): *Vom kritischen Ereignis zum Nutzungsproblem: Die qualitative Analyse in diagnostischen Usability Tests*. In: HASSENZAHL, M.; PEISSNER, M. (Hrsg.) (2004): *Usability Professionals 2004*. Stuttgart: GC-UPA e.V., pp. 142-146.

KRÄMER, N.C.; NITSCHKE, J. (2002). *Ausgabemodalitäten im Vergleich: Verändern sie das Eingabeverhalten der Benutzer?* In MARZI, R.; KARAVEZYRIS, V.; ERBE H.-H.; TIMPE, K.-P. (Hrsg.) (2002): *Bedienen und Verstehen. Tagungsband 4. Berliner Werkstatt Mensch-Maschine-Systeme*. Düsseldorf: VDI-Verlag, pp. 231-248.

KRÄMER, N.C.; BENTE, G. (2002): *Virtuelle Helfer: Embodied Conversational Agents in der Mensch-Computer-Interaktion*. In BENTE, G.; KRÄMER, N.C.; PETERSEN, A. (Hrsg.) (2002): *Virtuelle Realitäten*. Göttingen: Hogrefe Verlag für Psychologie, pp. 203-225.

KREITZBERG, Ch. (1996): *Managing for usability*. In: ALBER, A.F. (Hrsg.) (1996): *Multimedia: A Management Perspective*. Belmont, CA: Wadsworth, pp. 65-88

LIN, T.; OMATA, M.; HU, W.; IMAMIYA, A. (2005): *Do physiological data relate to traditional usability indexes?* In: *ACM International Conference Proceedings; Vol. 122*. Narrabundah: CHISIG, pp. 1-10.

PRÜMPER, J.; ANFT, M. (1993): *Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - ein Fallbeispiel*. In RÖDIGER, K-H. (Hrsg.): *Software-Ergonomie '93: Von der Benutzungsoberfläche zur Arbeitsgestaltung*. Stuttgart: Teubner, pp. 145-156.

RIST, Th.; ANDRE, E.; BALDES, St.; GEBHARD, P.; KLESEN, M.; KIPP, M.; RIST, P.; SCHMITT, M. (2004): *A review on the development of embodied conversational agents and their application fields*. In: PRENDINGER, H.; ISHIZUKA, M. (2004): *Life-Like Characters: Tools, Affective Functions, and Applications*. Berlin et al.: Springer-Verlag, pp. 377-404.

SCHIMMACK, U. (1997): *Das Berliner-Alltagssprachliche-Stimmungs-Inventar (BASTI). Ein Vorschlag zur kontentvaliden Erfassung von Stimmungen*. In: *Diagnostica 1997 / 2*. Göttingen: Hogrefe-Verlag, pp. 150-173.

STRICKER, A. (2003): *Darf's ein bisschen menschlicher sein? – Virtuelle Charaktere am Point-of-Sale*. In: LINDNER, Ch. (Hrsg.) (2003): *Avatare. Digitale Sprecher für Business und Marketing*. Berlin et al.: Springer-Verlag, pp. 167-183.

## Internetquellen

ERGO-ONLINE: *Gesellschaft Arbeit und Ergonomie - online e.V.* Internet, Verifizierungsdatum: 6.03.2006 <<http://www.sozialnetz-hessen.de/ca/pq/mdl>>

NIELSEN, J.: *Nielsen Norman Group. UseIt.Com*. Internet, Verifizierungsdatum 6.03.2006 <<http://www.useit.com/papers/heuristic/severityrating.html>>

THE WILD DIVINE PROJECT: *The Journey to Wild Divine*. Internet, Verifizierungsdatum 26.03.2006 <<http://www.wilddivine.com>>

UNIVERSITÄT TRIER: WIKILingua Computerlinguistik Universität Trier. Internet,  
Verifizierungsdatum: 6.03.2006

<[http://www.uni-trier.de/uni/fb2/ldv/ldv\\_wiki/index.php/Type-Token-Relation](http://www.uni-trier.de/uni/fb2/ldv/ldv_wiki/index.php/Type-Token-Relation)>

(1) WIKIMEDIA FOUNDATION INC.: Wikipedia. Die freie Enzyklopädie. Internet,  
Verifizierungsdatum: 6.03.2006 <<http://de.wikipedia.org/wiki/Rendern>>

(2) WIKIMEDIA FOUNDATION INC.: Wikipedia. Die freie Enzyklopädie. Internet,  
Verifizierungsdatum: 13.03.2006 <<http://de.wikipedia.org/wiki/Hawthorne-Effekt>>

## **Eigenständigkeitserklärung nach §26 Abs. 6 der Magisterprüfungsordnung**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle wörtlichen und sinngemäßen Zitate aus diesen Quellen sind entsprechend gekennzeichnet.

Hildesheim, den 5. April 2006

Matthias Görtz

# Anhang A – Usability-Heuristiken

---

## 1 - Sichtbarkeit des Systemstatus

Das System sollte die Benutzer ständig darüber informieren, was geschieht, und zwar durch eine angemessene Rückmeldung in einem vernünftigen zeitlichen Rahmen.

## 2 - Übereinstimmung zwischen dem System und der realen Welt

Das System sollte die Sprache der Benutzer sprechen, und zwar nicht mit systemorientierter Terminologie, sondern mit Worten, Phrasen und Konzepten, die den Benutzern vertraut sind. Dabei soll die natürliche und logische Reihenfolge eingehalten werden.

## 3 - Benutzerkontrolle und –freiheit

Benutzer wählen Systemfunktionen oft fälschlicherweise aus und benötigen einen Notausgang, um den unerwünschten Zustand wieder zu verlassen. Dazu dienen Undo- und Redo-Funktionen.

## 4 - Konsistenz und Standards

Benutzer sollten sich nicht fragen müssen, ob verschiedene Begriffe oder Aktionen dasselbe bedeuten. Deshalb sind Konventionen einzuhalten.

## 5 - Transparenz

Objekte, Optionen und Aktionen sollten sichtbar sein. Dem Benutzer sollte jederzeit bewußt sein, welche Aktionen, Objekte und Optionen ihm zur Verfügung stehen.

## 6 - Flexibilität und Effizienz der Benutzung

Häufig auftretende Aktionen sollten vom Benutzer angepaßt werden können, um Fortgeschrittenen eine schnellere Bedienung zu erlauben.

## 7 - Hilfe beim Erkennen, Diagnostizieren und Beheben von Fehlern

Fehlermeldungen sollten in natürlicher Sprache ausgedrückt werden (keine Fehlercodes), präzise das Problem beschreiben und konstruktiv eine Lösung vorschlagen.

## 8 - Hilfe und Dokumentation

Jede Information der Hilfe oder Dokumentation sollte leicht zu finden sein, auf die Aufgabe abgestimmt sein und die konkreten Schritte zur Lösung auflisten. Außerdem sollte sie nicht zu lang sein.

## 9 – Motivation und Unterstützung

Das System sollte den Benutzer bei seiner Aufgabe unterstützen und motivieren. Dazu zählt die Anregung zur Nutzung genau so wie die Unterstützung während und Motivation nach erfolgreicher Aufgabenerledigung.

## Anhang B – Mängelliste der Experten-Evaluierung

Beschreibung der Mangellerscheinung	Kategorie	Bewertung
Der Befehl "Radio" führt direkt zu Unterauswahl ohne dass das Gerät vorher angezeigt wurde und somit bekannt ist welcher Sender aktiv ist oder welche Sender zur Auswahl stehen.	1 - Sichtbarkeit Systemstatus	3 - Bedeutendes Problem
Die lange Verzögerung zwischen Knopfdruck PTT und Systembereitschaft stellt eine große Gefahr für den Dialogfluss und die fehlerfreie Eingabe des Users dar.	1 - Sichtbarkeit Systemstatus	3 - Bedeutendes Problem
CD: Beschriftung "Lied <1..30>/CD <1..12>" irreführend: falsche Reihenfolge und kryptische Darstellung verwirren bei kurzem Blick auf Auswahl den Nutzer.	2 - Übereinstimmung System - Reale Welt	3 - Bedeutendes Problem
NAVI: Der Befehl "Suche um Stadt" / "Suche in der Nähe" und ihre Funktionen sind nicht ersichtlich.	2 - Übereinstimmung System - Reale Welt	3 - Bedeutendes Problem
Radiosender kann man nicht mit Namen aufrufen. ("Jump", "NDR2")	2 - Übereinstimmung System - Reale Welt	3 - Bedeutendes Problem
TEL: "Und weiter?" ist eine zu direkte Frage, die den Benutzer dazu auffordert eine Eingabe fortzusetzen, obwohl nicht offensichtlich ist, dass die Eingabe nicht schon beendet sein könnte. Eine offener Frage sollte gestellt werden.	3 - Benutzerkontrolle- und Freiheit	3 - Bedeutendes Problem
CD: "Anspielen" und "Mix an/aus" stehen zusammen sind aber inhaltlich keine zusammengehörigen Funktionen.	4 - Konsistenz und Standards	3 - Bedeutendes Problem
TEL: Das System gibt hier Feedback "Die Nr. wird gewählt", im CD-Wechsler fehlt allerdings solche eine Bestätigung wie z.B. "CD6 ausgewählt." Dies stört die Konsistenz.	4 - Konsistenz und Standards	3 - Bedeutendes Problem
Terminologie: "CD-Wechsler" / Beschriftung: "CD-Audio" nicht konsistent	4 - Konsistenz und Standards	3 - Bedeutendes Problem
Das Vorgelesene und das Visuelle stimmen nicht überein.	4 - Konsistenz und Standards	3 - Bedeutendes Problem
Befehlsliste ist unübersichtlich und nicht selbsterklärend.	5 - Transparenz	3 - Bedeutendes Problem
Bei "buchstabieren" fordert das System auf: "Bitte geben Sie ein Sprachkommando ein."	5 - Transparenz	3 - Bedeutendes Problem

Beschreibung der Manglerscheinung	Kategorie	Bewertung
"Globale Sprachkommandos" stehen nicht überall zur Verfügung. In Unterebenen der Applikationen ist ein Wechsel zu anderen Applikationen nicht möglich.	5 - Transparenz	3 - Bedeutendes Problem
Bei Telefonnummereingabe ist nicht erkenntlich welches Format und wieviele Nummern eingegeben werden können.	5 - Transparenz	3 - Bedeutendes Problem
NAV: "Eingestellte Optionen" wird nicht als möglicher Befehl nach Kartenansicht in Befehlsliste geführt, ist aber verfügbar.	5 - Transparenz	3 - Bedeutendes Problem
Abbruch ist nicht erwartungskonform. Der Benutzer befindet sich weiterhin in Applikation und Kontext-Menü. (Es erfolgt kein Sprung in oberste Ebene)	5 - Transparenz	3 - Bedeutendes Problem
Bei erneuter Straßeneingabe ist nicht mehr erkennbar, welche die vom System erkannte Stadt ist.	5 - Transparenz	3 - Bedeutendes Problem
Starre Kommandostruktur verhindert schnellen Durchlauf durch Eingabesequenz.	6 - Flexibilität und Effizienz	3 - Bedeutendes Problem
Buchstabierende Eingabe nicht von Beginn an möglich.	6 - Flexibilität und Effizienz	3 - Bedeutendes Problem
Wiederholt man bei Stadtbestätigung die Eingabe, so verwendet das System die dann verstandene Stadt und leitet ohne Bestätigungsabfrage zu "falscher Stadt"	7 - Hilfe: Erkennen, Diagnostizieren und Korrigieren	3 - Bedeutendes Problem
Hilfe ist nicht sehr hilfreich, da nur die Menüoptionen vorgelesen werden. Die Erwartung des Benutzers ist eine detaillierte Erklärung der einzelnen Funktionen und ihrer Bedeutung bzw. eine Anleitung im Sinne von "Was möchten Sie tun?" oder eine Erläuterung der möglichen Eingabeformate etc.	8 - Hilfe und Dokumentation	3 - Bedeutendes Problem
Befehl "neues Ziel" während aktiver Zielführung bricht zwar Zielführung ab, ermöglicht aber keine direkte neue Zieleingabe.	Bug	3 - Bedeutendes Problem

Beschreibung der Mangelercheinung	Kategorie	Bewertung
CD: Rückmeldung/Bestätigung des Systems fehlt. Der Benutzer weiss nicht, ob das System ihn richtig verstanden hat und welche CD nun gespielt wird.	1 - Sichtbarkeit Systemstatus	2 - Geringfügiges Problem
Standby-Modus der Interface Agentin impliziert Aktion bzw. Sprachausgabe, dabei ist "Helen" gerade inaktiv. Nicht erwartungskonform.	1 - Sichtbarkeit Systemstatus	2 - Geringfügiges Problem
RADIO: Begriff "AutoStore" ist nicht selbsterklärend.	2 - Übereinstimmung System - Reale Welt	2 - Geringfügiges Problem
NAVI: "Zielführung starten" kein dem Nutzer intuitiv bekannter Ausdruck.	2 - Übereinstimmung System - Reale Welt	2 - Geringfügiges Problem
Semantik "Anspielen" / "Start" nicht eindeutig.	2 - Übereinstimmung System - Reale Welt	2 - Geringfügiges Problem
Terminologie: "AutoStore", "Zielführung", "vorheriges" sind dem Benutzer nicht selbsterklärend.	2 - Übereinstimmung System - Reale Welt	2 - Geringfügiges Problem
Terminologie: "CD-Wechsler" "CD-Audio" "CD-Player" "CD-Changer" unverständlich.	2 - Übereinstimmung System - Reale Welt	2 - Geringfügiges Problem
Eingabesequenz wird nicht automatisch an letzter Stelle wieder aufgenommen.	3 - Benutzerkontrolle- und Freiheit	2 - Geringfügiges Problem
Nach Applikationswechsel erwartet das System sofort eine weitere Eingabe ohne dem Nutzer die Auswahlmöglichkeiten anzugeben.	3 - Benutzerkontrolle- und Freiheit	2 - Geringfügiges Problem
NAVI: Bei Unterbrechung der Zielführung fragt das System nach "Soll die aktive Zielführung beendet werden?" Diese Rückbestätigung findet im System bei anderen Eingaben nicht statt und führt zu Inkonsistenz.	4 - Konsistenz und Standards	2 - Geringfügiges Problem
NAVI: Der Befehl "Suche um Stadt" ist im gesprochenen Text "Suche um größere Stadt".	4 - Konsistenz und Standards	2 - Geringfügiges Problem
Terminologie: "Abbruch" ist doppelt belegt. A) Abbruch der Ansage B) Abbruch der Eingabesequenz	4 - Konsistenz und Standards	2 - Geringfügiges Problem
Gestik des Avatars ist nicht einheitlich / inkonsistent.	4 - Konsistenz und Standards	2 - Geringfügiges Problem



Beschreibung der Manglerscheinung	Kategorie	Bewertung
Welche Kommandos werden verstanden? Grad der "Synonymisierung"?	5 - Transparenz	2 - Geringfügiges Problem
Telefonnummereingabe: "+" wird als Option angezeigt ist aber nur im Zusammenklang mit weiterer Zahl im System implementiert.	5 - Transparenz	2 - Geringfügiges Problem
Bei Nachfrage in der Stadteingabe ("Sagten Sie XYZ?") werden Optionen "Ja"/"Nein" nicht in Befehlsliste angezeigt.	5 - Transparenz	2 - Geringfügiges Problem
TEL: Korrektur nach langen Ziffereingaben bedeutet lange Korrekturzeiten, da alle Ziffern noch einmal wiederholt werden müssen. Eine Korrektur einzelner Digits sollte möglich sein.	7 - Hilfe: Erkennen, Diagnostizieren und Korrigieren	2 - Geringfügiges Problem
Wird die Eingabe nicht verstanden, so ist keine Zweiteingabe möglich, sondern nur "Buchstabieren" / Zeilennummer der Alternativen etc..	7 - Hilfe: Erkennen, Diagnostizieren und Korrigieren	2 - Geringfügiges Problem
Einführung bedient sich keiner Visualisierung und führt keine Beispiele vor. Eine Anleitung zur Bedienung ist es somit nicht.	8 - Hilfe und Dokumentation	2 - Geringfügiges Problem
RADIO: Bei FM Wechsel erfolgt zwar ein FM Wechsel, der Sender bleibt aber gleich auch wenn er in anderer FM Ebene nicht gespeichert ist.	Bug	2 - Geringfügiges Problem
Ansage kann mit PTT-Knopf nicht unterbrochen werden.	Bug	2 - Geringfügiges Problem
"Alles löschen" wird in der Telefonnummereingabe nicht verstanden, sondern nur "löschen".	Bug	2 - Geringfügiges Problem

Beschreibung der Mangelercheinung	Kategorie	Bewertung
Terminologie: "Applikation" nicht erwartungskonform - es handelt sich um "Geräte".	2 - Übereinstimmung System - Reale Welt	1 - Kosmetisches Problem
Terminologie: "Zieleingabe" unnatürliche Sprache	2 - Übereinstimmung System - Reale Welt	1 - Kosmetisches Problem
Mehrstellige Ziffern können nicht eingegeben werden. (Z.Bsp. "Dreiundfünfzig Elf drei")	2 - Übereinstimmung System - Reale Welt	1 - Kosmetisches Problem
Terminologie: Anwendung heisst "SprachDialogSystem" nennt sich aber "Sprachbediensystem".	4 - Konsistenz und Standards	1 - Kosmetisches Problem
Shortcuts stehen zwar anscheinend zur Verfügung sind aber nicht erkenntlich.	5 - Transparenz	1 - Kosmetisches Problem
CD: Der Hilfetext ist sehr lang und kann nicht unterbrochen werden. Zudem kann der Benutzer die Hilfe nicht auf einen speziellen Punkt anwenden ("Was bedeutet xyz?")	8 - Hilfe und Dokumentation	1 - Kosmetisches Problem
"Sie sind im Telefon/Radio...." ungünstig gewählte Ausdrucksweise	2 - Übereinstimmung System - Reale Welt	0 - Kein Usability Problem
RADIO: Keine Eingabe einer Frequenz ("101 komma 4") möglich.	2 - Übereinstimmung System - Reale Welt	0 - Kein Usability Problem
Kaum Rückbestätigung/Verifizierung abgefragt. Dies kann große Auswirkungen bei falsch verstandener Eingabe haben. (s.Bsp. Lässt sich Wahlwiederholung nicht abbrechen)	3 - Benutzerkontrolle- und Freiheit	0 - Kein Usability Problem
Dem Nutzer wird zwar mitgeteilt, dass die "Meine Ziele"-Liste leer ist, es wird aber nicht konstruktiv erläutert, wie er dort Ziele abspeichern kann.	8 - Hilfe und Dokumentation	0 - Kein Usability Problem
Beschränkte Kommandosprache hindert Benutzer an Zugriff auf alle Systemfunktionalitäten.	9 - Motivation und Unterstützung	0 - Kein Usability Problem
Eingabe der Stadt "Neustadt" -> es erfolgt keine Klärung der Mehrdeutigkeit.	Bug	0 - Kein Usability Problem

# Anhang C – Aufgabenszenarien des Benutzertests

---

## EINFÜHRUNG

---

### Herzlich Willkommen zum Benutzertest „Innovative Fahrerinformationssysteme“ der Universität Hildesheim und der Blaupunkt GmbH.

Heute werden Sie die Möglichkeit haben, zwei Prototypen von Sprachdialogsystemen kennzulernen. Diese Systeme sind wie folgt aufgebaut:

1. Der Benutzer drückt den Push-To-Talk Knopf (*linker Mausknopf*)
2. Es erfolgt ein Signalton
3. Das Mikrofonsymbol zeigt an, dass man einen Sprachbefehl sprechen kann
4. Der Benutzer spricht das gewünschte Sprachkommando

Die zur Verfügung stehenden Sprachbefehle erscheinen dabei auf dem Display. Nehmen Sie sich ruhig im Vorfeld die Zeit die Befehlsliste erst einmal gut durchzulesen. Ist später dann das Mikrofon bereit, können Sie das gewünschte Kommando sprechen. Sie können sich auch jederzeit mit dem Befehl „Hilfe“ noch einmal vorlesen lassen, was für Funktionen Ihnen zur Verfügung stehen.

Die Applikationen:

Der <i>CD-Wechsler</i>	-	CDs und Liednummern können ausgewählt und symbolisch abgespielt werden
Das <i>Radio</i>	-	Sender lassen sich anhören, suchen und abspeichern
Das <i>Telefon</i>	-	Sie können Nummern diktieren und wählen lassen
Die <i>Navigation</i>	-	Sie können hier Ihr Fahrtziel eingeben und sich dann den Weg auf dem Display symbolisch anzeigen lassen. Dabei gilt folgende Reihenfolge: <i>Stadt &gt; Strasse &gt; Hausnummer</i>

Nach der Eingabe in das System werden Sie an wichtigen Stellen gebeten, die Eingabe zu bestätigen. Hat das System Sie falsch verstanden besteht die Möglichkeit, aus einer Liste die passende Eingabe auszuwählen oder die Eingabe nachfolgend zu buchstabieren.

Im Folgenden haben Sie nun die Zeit das System etwas genauer zu erforschen. Anschließend bekommen Sie einige kurze Aufgaben. Haben Sie diese mit dem System erfüllt, können Sie in einem Fragebogen Ihre Eindrücke wiedergeben.

*Vielen Dank für Ihre Mithilfe und viel Spaß!*

---

## AUFGABENBLATT 1

---

*Willkommen zu Ihrer ersten Probefahrt. Stellen Sie sich vor, es ist ein Februartag im Jahr 2010. Sie steigen in einen Mietwagen, um zu einem wichtigen Geschäftstermin zu fahren. Bevor Sie die Fahrt antreten, richten Sie sich in dem neuen Auto ein: Sie probieren das Fahrerinformationssystem mal ein wenig aus.*

*Nehmen Sie sich ruhig ein paar Minuten Zeit, um die Sprachbedienung, das Radio und das Navigationssystem kennenzulernen.*

*Bevor Sie die Fahrt nun antreten, möchten Sie noch einiges erledigen:*

*Radio*

---

- ☐ Sie möchten sich den nächsten Radiosender anhören.
- ☐ Sie suchen sich einen Sender der Ihnen gefällt und speichern ihn auf Speicherplatz 5.
- ☐ Sie lassen sich nun vom Radio automatisch die aktuell verfügbaren Sender auf den Speicherplätzen ablegen.
- ☐ Abschließend wechseln Sie zu SWR3.

*Navigation*

---

Vor Antritt Ihrer Fahrt, möchten Sie nun im Navigationssystem Ihre Zieladresse eingeben. Sie haben heute einen Termin bei

		<b>Robert Bosch GmbH</b>
Rainer Meinholz		
Dipl. Ing.		
		Robert Bosch GmbH
		Entwicklung
		Kirchstr. 113
		70173 Stuttgart
Email: <a href="mailto:rainer.meinholz@de.bosch.com">rainer.meinholz@de.bosch.com</a>	Telfon : 0711-917254	
	Mobil : 0172-3865433	

- ☐ Sie wechseln zunächst zum Navigationssystem.
- ☐ Dann lassen Sie sich vom System die Funktionsweise der Navigation erläutern
- ☐ Nun können Sie das Navigationssystem auf die Adresse von Herrn Meinholz programmieren.
- ☐ Abschließend möchten Sie sich zusätzlich zu den Pfeilsymbolen noch die Karte mit anzeigen lassen.

*Gute Fahrt!*

---

## AUFGABENBLATT 2

---

*Ihr Geschäftstermin ist gut gelaufen und Sie können noch heute die Rückfahrt antreten. Sie bekommen wieder einen Mietwagen zur Verfügung gestellt, um nach Hause fahren zu können. Auch hier schauen Sie sich CD-Wechsler und das Telefon erst genauer an.*

*Bevor Sie die Fahrt nun antreten, möchten Sie nun noch einiges erledigen:*

---

### CD-Wechsler

- ☐ Sie schalten zum CD-Wechsler und wählen CD 5 aus.
- ☐ Das erste Lied gefällt Ihnen nicht und Sie lassen sich automatisch jedes Lied auf der CD kurz anspielen.

Ihnen fällt ein, dass Sie ihre Lieblings-CD dabei haben.

- ☐ Sie starten das 12. Lied auf CD 7.

---

### Telefon

Es fällt Ihnen noch ein, dass Sie vergessen hatten Herrn Meinholz etwas zu sagen.

- ☐ Sie wechseln zum Telefon.
- ☐ Sie rufen ihn an seinem Arbeitsplatz an.
- ☐ Er antwortet leider nicht. Sie legen auf und versuchen es auf seinem Handy.
- ☐ Es ist besetzt. Sie versuchen es noch einmal.

---

### Navigation

Nachdem nun alles in Stuttgart erledigt ist, können Sie sich auf den Heimweg machen.

- ☐ Programmieren Sie das Navigationssystem auf eine Stadt Ihrer Wahl. Verzichten Sie dabei auf die Straßeneingabe.

*Nun können Sie Ihre Heimfahrt beginnen!  
Vielen Dank!*

## Anhang D – Ergebnisse der Fragebögen

---

**Item 3:** *Was gefiel Ihnen an diesem sprachbedienbaren Fahrerinformationssystem?*

### SDS Demonstrator

Weniger Ablenkung beim Fahren
Intuitive Bedienung. Gute Spracherkennung. Übersichtliches Menü.
Überraschend klar zu bedienen; klare Vorteile vor allem bei Texteingabe (navigation:strasse) gegenüber konventionellen Geräten
Es ist sehr einfach; man braucht keine Vorkenntnisse
Überschaubar, keine unnötigen Optionen
Die einfache Bedienung über einige Befehle ohne längere Zeit durch "Knöpfe drücken" a.Ä. abgelenkt zu werden.
Umfangreiche, schnelle Möglichkeit an viele Informationen zu kommen. Geringer Lernaufwand für mich persönlich
Ziemlich einfach zu bedienen; gute Spracherkennung
Abgesehen von Zahlen (Hausnr.) gute Spracherkennung. Wenige Befehle nötig. Mögliche Befehle werden angezeigt.
Gute Spracherkennung
Die Eingewöhnungsphase ist relativ kurz. Übersichtlicher Aufbau des Displays. Man erkennt relativ schnell seine Fehler und kann sie problemlos korrigieren.
Angenehme Stimme
Alle Infos auf einem Display
Keine zusätzliche Ablenkung durch "Figur"
Klare Benutzerführung ohne visuelle Ablenkung durch Helen (ihr Sinn im ersten System ist mir immer noch nicht klar, sie trug nicht zur Verständlichkeit des Systems bei)
CD-Wechsler und Navi
Die Hilfe ist nur durch Sprache verfügbar, es gibt keine störende Frau mehr im Bild
Keine irritierende Figur, ansonsten s. System 1
Die angenehme Stimme. Das klare und übersichtliche GUI.
Einfache Bedienung bei Navigation.
Die Spracheingabe selbst. Die Bündelung weiterer Funktionen wie Tel/CD. Das Fehlen eines Agenten hat weniger von den Funktionen abgelenkt.
Es ist unkompliziert
Wenig Ablenkung durch "rumfummeln" an Knöpfen

## Helen

Großes zentrales Display, parallel wäre ein Touchscreen nicht schlecht
die Sprachbedienung / die Symbole
Anzeige der möglichen Befehle in jedem Einzelsystem / Systemschritt. Einzelne Systemelemente eindeutig getrennt. Status/Modus des Systems.
Geringer Lernaufwand. Man findet sich schnell zurecht. System versteht Sprache relativ gut, wenn man gewisse Dinge (bspw. Intonation) beachtet
Die dynamische Menüführung. Ich hatte nicht das Gefühl, mich nur in einer Kategorie z.b. Navi zu befinden, sondern jederzeit wechseln zu können. Die Spracherkennung ist sehr ausgereift.
Nach Eingewöhnungszeit sicher guter Beitrag zu Verkehrssicherheit und Fahrkomfort
Die Sprachausgabe ist relativ angenehm.
Spracherkennung ok
Einfache Bedienung, gutes Sprachverständnis
Das es überhaupt sprachgesteuert ist. Dass alle wesentlichen Funktionen für eine Fahrplanung enthalten sind.
Faszination, da ich bisher nicht damit konfrontiert wurde.
Vereint viele Funktionen; kompakt
Idee der Sprachbedienung an sich gut
weniger Ablenkung beim Fahren
s. System 1 / freundlicher Avatar
Die aufmerksame Dame, vermute aber, dass sie über kurz oder lang nervt
Daß man sich mit den Sprachbefehlen voll und ganz auf den Verkehr konzentrieren kann.
Die einfach Bedienung durch übersichtliche Anzahl von Sprachbefehlen
Leicht zu bedienen; gute Spracherkennung
Intuitiv steuerbar. Schnell zu erlernen.
Gute Korrekturmöglichkeit
Je länger man sich damit beschäftigt und je besser man sich damit auskennt, desto mehr Spass macht es.
Einige wichtige Optionen / Befehle werden nochmal benannt
Kein Unterschied zu system 1

**Item 3:**

*Was gefiel Ihnen nicht an diesem sprachbedienbaren Fahrerinformationssystem?*

SDS Demonstrator

Nicht immer verstanden. Die Informationen werden nicht immer verstanden.
Keine Shortcuts; dass der Piepton abgewartet werden muss; dass man bei Fehlern direkt im Hauptmenü landet; Unklarheiten bei Rückfragen (meinten sie stuttgart?) da Antworten nicht vorgegeben
Die Eingabe der Straße ist umständlich, da die Aussprache bei jedem Menschen anders ist
Die Anweisungen setzen zum Teil Vorkenntnisse voraus. Zahlenangaben werden nur schlecht verstanden.
Warten auf den Piepton, wenn Antwort schon bekannt ist. Befehle zu lang, könnte kürzer sein.
Bei manchen Befehlen war nicht ganz klar, welche Handlungsoptionen zur Verfügung stehen ("meinten sie stuttgart?")
Z.T. wurde Eingabe nicht verstanden.
Man musste extrem deutlich sprechen (hohe Konzentration / Ablenkung). Kann mir vorstellen, dass für einige die Begriffe zu kompliziert sind (Lernaufwand)
Begriffe nicht einheitlich (AutoStore engl.)
Fehlererkennung von Sprache. Langwierige Eingaben. Wenn das System angelernet werden muss.
Wenig Rückmeldung
Mangelnde Spracherkennung. Man gerät ins Stocken, fühlt sich hilflos und etwas verunsichert (in manchen Fällen).
Etwas unübersichtlich von der Optionendarstellung / Struktur. Die verwendeten Begriffe sind nicht intuitive verstehbar und erschließen sich nicht auf den ersten Blick.
Mich würde es ablenken
Ewige Wiederholung von Straßen usw
Zeitaufwand. Vorgegebene Wortwahl
Zu viele und zu lange Schritte um zu Ziel zu kommen. Telefonbedienung extrem schlecht. Bin auf Tastatur 10x schneller. Wiederholung der Zahl nicht nötig. Wahlwiederholung sollte automatisch ausgeführt werden.
Offensichtlich Schwierigkeiten beim Verstehen der Telefonnummer
TelNrEingabe zu langatmig. Piepton nichts für spontane Schnellsprecher
Dass es meine Rufnr nicht versteht und ich nicht weiss, wie ich die Zahlen aussprechen soll, damit es mich versteht.



Nach wie vor ist der Signalton nach dem PTT zu spät. Anstelle von Lied (x) würde mir Track (x) besser gefallen, aber das ist eine Gewohnheitsfrage.
Auch wieder warten bis Mikro frei. Befehlsbezeichnungen wieder manchmal nicht so deutlich. TelNrEingabe nur blockweise (muss man erst drauf kommen)
Keine klare Anweisung bei Telefonnummerneingabe. Zu starr.
Die Telefonfunktion ist langwierig. (einzelne Zahlen eingeben)
Die Eingabe von einzelnen Nummern anstelle von mehrstelligen Nummern. Die Eingabe war zu langwierig.
Bei Eingabe der Telefonnummer zu hoher Aufwand
Resistenz bei falschen Straßennamen; Kartendisplay zu klein

#### Helen

Die Notwendigkeit sich genau an die Sprachanweisung zu halten. Nicht verstehen des Ortsnamens.
Zu aufwendig. Bsp: Um bei meinem jetzigen System zum nächsten Sender zu wechseln ist ein Tastendruck nötig, hier hingegen Tastendruck und Sprachanweisung. "Ohne Moderatorin die nervt"
Gewöhnungsbedürftig / erfordert deutliche Aussprache
Assistentin etwas zu aktiv, präsent. Piepton muss bis zur nächsten Eingabe abgewartet werden.
In den jew. Menüs müsste auch stehen, dass man in die anderen wechseln kann. Sonst sucht man nach den Wörtern und weiß nicht, wie man in die anderen Menüs wechselt.
Die "Hilfsperson" stand teilweise mitten im Bild. Generell würde ich diese nur durch Sprache ersetzen wollen. Das Signal nach drücken des PTT-Buttons müsste für meinen Geschmack schneller ertönen.
Dauert lange bis man reden kann. Die Tante stört etwas, einmal stand sie sogar vor der Liste. Listeneinträge nicht eindeutig. Funktionserklärung wäre wichtig.
Die Hausnummerneingabe
Anfangs verwirrend, nicht eindeutige Anwendungshinweise
Manchmal keine Information über die nächsten Schritte oder Fehlfunktion. Lange Befehlseingaben nötig (kurze Befehle wären besser)
Manch Begrifflichkeiten waren unklar. Z.B. AutoStore für die Programmierung der Sender. Es war nicht immer klar über welche Tasten (oder Befehle) ich meine Aufgabe oder mein Ziel erreichen konnte.
Bei Erstbenutzung unübersichtlich; die begriffe und Anordnung unbekannt
Die "dame", die da rumhampelt, irritiert mich. Sie hilft nicht weiter, ist nicht schön anzusehen & bewegt ihre Lippen nicht zur Sprache; Telnr. Eingabe

**Item 6:**

*Welches der beiden Systeme würden Sie für den Einsatz in Ihrem Fahrzeug bevorzugen? Warum?*

SDS

Frau im Display überflüssig.
Ohne Figur
Beide Systeme sind vom Umfang/Komfort gleich, allerdings bringt mir die persönliche Hilfe keinen weiteren Vorteil (visuelle Ablenkung). System 2 ist weniger "aufdringlich".
Die Hilfe mich beim 1. System stört.
Keine Figur.
Ich benötige die Frau zur Ansprache nicht. Sie lenkt eher vom Straßenverkehr ab.
Ich fand den Avatar zu ablenkend.
Ich würde nicht durch die Dame abgelenkt
Scheinbar für mich persönlich einfacher zu bedienen
Wegen der Frau im Bild
Keine Unterschiede ersichtlich bis auf "Empfangsdame" - keine Vorteile; wird wahrscheinlich nerven
Weil mich die Frau eher gestört hat; sie war nicht hilfreich
Animation Spielerei, total unnötig
Animierte Frau ist zwar witzig, aber nicht unbedingt nötig; kann mir vorstellen, dass sie nach längerer Zeit eher lästig wird.
Die Frau im System 2 stört und lenkt ab.
Ohne Frau (Person) gefällt es mir besser.
Sprecherin ist Schnick-Schnack. Ich gucke eh nicht hin wenn ich z.B. den Wechsler steuere.
Keine Ablenkung durch Spielerei-Avatar auf Bildschirm
Figur ist nicht nötig
Beim Auto fahren kann man ja nicht auf den Bildschirm und die Frau angucken

Helen

Weil mir das Telefon nicht gefällt.
Der Avatar blockiert zwar den Blick auf das GUI, stellt aber einen "persönlichen Kontakt" zum System her.
Klare Anweisung. Freundlicheres Display
Einfach in der Bedienung, Ansprechendes Display
Bei der Fahrt ist der Avatar egal, bei der Routenplanung vor der Fahrt oder bei Stops ist der Avatar aber angenehm. Während der Fahrt würde ich das System nicht betrachten.
Die Dame gefällt mir, muß man aber bei Bedarf ausblenden können.
Es wird ein bißchen mehr erklärt und Feedback gegeben. Viel genutzte / wichtige Optionen werden sprachlich noch einmal benannt.

## Anhang E – Auswertung physiologischer Kennzahlen





